

การเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สิน
ของประชากรไทยโดยวิธีเคมีนและดีบีสแกน

COMPARISON OF HOUSEHOLD SEGMENTATION BY DEBT OF
THE THAI POPULATION BY K-MEANS AND DBSCAN METHODS



ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ภาควิชาสถิติ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2565

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COMPARISON OF HOUSEHOLD SEGMENTATION BY DEBT OF
THE THAI POPULATION BY K-MEANS AND DBSCAN METHODS



Firdao Laesor
Wilairat Saysaard
Sirithip Thongto

A SPECIAL PROBLEM SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS)
DEPARTMENT OF STATISTICS, SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
ACADEMIC YEAR 2022

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ

การเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของ
ประชากรไทยโดยวิธีเคมีนและดีบีเอสแกน

Comparison of household segmentation by debt of
the Thai population by K-means and DBSCAN
methods

ชื่อนักศึกษา

นางสาวพิรดาว แลซอ รหัสนักศึกษา 62050807

นางสาววิไลรัตน์ สายสะอาด รหัสนักศึกษา 62050834

นางสาวสิริทิพย์ ทองโต รหัสนักศึกษา 62050843

ปริญญา

วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)

ภาควิชา

สถิติ

ปีการศึกษา

2565

อาจารย์ที่ปรึกษา

ผศ.ดร.ยุวดี กล่อมวิเศษ

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้
ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (สถิติประยุกต์)
ประจำปีการศึกษา 2565

คณะกรรมการสอบ	ลายมือชื่อ
ผศ.ดร.กนกวรรณ ลีโรจนประภา ประธานกรรมการ	
ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ กรรมการ	
ผศ.ดร.ยุวดี กล่อมวิเศษ กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ	การเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของ ประชากรไทยโดยวิธีเคมินและดีปัสแกน	
ชื่อนักศึกษา	นางสาวพิรดาว แลขอ	รหัสนักศึกษา 62050807
	นางสาววิไลรัตน์ สายสะอาด	รหัสนักศึกษา 62050834
	นางสาวสิริทิพย์ ทองโต	รหัสนักศึกษา 62050843
ปริญญา	วิทยาศาสตร์บัณฑิต (สถิติประยุกต์)	
ภาควิชา	สถิติ	
ปีการศึกษา	2565	
อาจารย์ที่ปรึกษา	ผศ.ดร.ยุวดี กล่อมวิเศษ	

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของคนไทยโดย
ใช้วิธีเคมินและดีปัสแกน เพื่อออกแบบแดชบอร์ดสำหรับใช้ในการนำเสนอข้อมูลภาระหนี้สินของคนไทย
โดยนำข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือนปี พ.ศ. 2564 ซึ่งเป็นข้อมูลทุติยภูมิ
รวบรวมโดยสำนักงานสถิติแห่งชาติจากแบบสำรวจครัวเรือนส่วนบุคคลทั้งในเขตและนอกเขตเทศบาลทั่ว
ประเทศ จำนวน 46,840 ครัวเรือน แบ่งเป็นครัวเรือนที่มีหนี้สิน 24,238 ครัวเรือน และครัวเรือนที่ไม่มี
หนี้สิน 22,602 ครัวเรือน ในกลุ่มครัวเรือนที่มีภาระหนี้สิน จำนวนกลุ่มที่เหมาะสมสำหรับวิธีเคมินโดย
พิจารณาจาก Silhouette score คือ 2 กลุ่ม และการแบ่งกลุ่มด้วยวิธีดีปัสแกน พบว่าไม่สามารถแบ่งกลุ่ม
ได้ จากนั้นวัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี Silhouette coefficient พบว่าการแบ่งกลุ่มของข้อมูลใน
กลุ่มครัวเรือนที่มีภาระหนี้สินวิธีที่ดีที่สุดคือวิธีเคมินที่จำนวนกลุ่มเท่ากับ 2 (score = 0.22) และกลุ่ม
ครัวเรือนไม่มีหนี้สิน จำนวนกลุ่มที่เหมาะสมสำหรับวิธีเคมินโดยพิจารณาจาก Silhouette Score คือ 2
กลุ่ม และการแบ่งกลุ่มด้วยวิธีดีปัสแกนได้จำนวนกลุ่มที่เหมาะสมจากวิธี Nearest Neighbor คือ 2 กลุ่ม
จากนั้นวัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี Silhouette coefficient พบว่าการแบ่งกลุ่มของข้อมูลใน
กลุ่ม ครัวเรือนที่ไม่มีภาระหนี้สินวิธีที่ดีที่สุดคือวิธีเคมินที่จำนวนกลุ่มเท่ากับ 2 (score = 0.35) จากนั้นทำ
การทดสอบสมมติฐานความแตกต่างของค่าเฉลี่ยระหว่าง Cluster ในกลุ่มครัวเรือนที่มีภาระหนี้สิน และ
กลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน พบว่า ตัวแปรที่มีค่าเฉลี่ยแตกต่างกันระหว่าง Cluster ทั้งกลุ่มครัวเรือนมี
ภาระหนี้สิน และกลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน มีหลายตัวแปร เช่น จำนวนสมาชิกในครัวเรือน จำนวน
ผู้รายได้ของครัวเรือน จำนวนผู้พิการในครัวเรือน จำนวนผู้ที่กำลังศึกษาในครัวเรือน จำนวนสวัสดิการด้าน
การรักษาพยาบาล จำนวนโครงการช่วยเหลือจากรัฐบาล เป็นต้น จากนั้นผู้วิจัยได้นำตัวแปรที่ค่าเฉลี่ยไม่
แตกต่างกันระหว่าง Cluster ออกจากการแบ่งกลุ่มแล้วได้ทำการแบ่งกลุ่มอีกครั้ง พบว่าได้ค่า Silhouette

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของงานเพื่อการศึกษาเท่านั้น เมื่อผู้ยู่ใช้ต้องระบุชื่อหน่วยงานที่
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

coefficient เพิ่มขึ้นเป็น 0.235 และ 0.390 สำหรับกลุ่มครัวเรือนที่มีภาระหนี้และไม่มีภาระหนี้สิน ตามลำดับ ทั้งนี้ผู้วิจัยหวังเป็นอย่างยิ่งว่าหน่วยงานที่เกี่ยวข้องทั้งในภาครัฐและเอกชน จะสามารถใช้งานวิจัยนี้เป็นแนวทางในการแก้ไขปัญหาภาระหนี้สินของประชากรไทย และสามารถช่วยลดความเสี่ยงในการก่อภาระหนี้สินของครัวเรือนไทยในลำดับต่อไป

คำสำคัญ : เคมีน, ดีปีสแกน, หนี้ของครัวเรือน, การแบ่งกลุ่ม, Silhouette score



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Comparison of household segmentation by debt of the Thai population by K-means and DBSCAN methods	
Student	Miss Firadao Laesor	Student ID 62050807
	Miss Wilairat Saysiaard	Student ID 62050834
	Miss Sirithip Thongto	Student ID 62050843
Degree	Bachelor of Science (Applied Statistics)	
Department	Statistics	
School	Science	
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)	
Academic Year	2022	
Advisor	Asst. Prof. Dr. Yuwadee Klomwises	

Abstract

This study aims to compare clustering techniques that use to investigate the debt of the Thai population based on the K-means and DBSCAN methods and to design dashboard which visualize cluster of Thai population debt. The data used in this study is from the 2021 survey on the economic and social status of households conducted by the National Statistical Office, which collected data from 46,840 households, including 24,238 households with debt and 22,602 households without debt, both within and outside municipal areas throughout the country. For indebted households, optimal number of cluster is 2 clusters for K-means based on the Silhouette Score, while the DBSCAN method was unable to perform data clustering. Then, the performance of the clustering methods was measured using the Silhouette coefficient and it was found that the K-means clustering method with 2 clusters (score = 0.22) was the best for making data clustering for indebted households. For debt-free households, optimal number of cluster is 2 clusters for K-means based on the Silhouette score and the DBSCAN method has 2 clusters as optimal number of cluster based on the Nearest Neighbor method. Then, the performance of the clustering methods was measured using the Silhouette coefficient and it was found that the K-means clustering method with 2 clusters (score = 0.35) was the

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

best for making data clustering for debt-free households. Afterwards, the hypothesis testing of differences in cluster means between clusters of indebted households and those of debt-free households was conducted. It found that there were several variables with different mean values between the two clusters of indebted households and debt-free households, such as the number of household members, number of individuals with income in the household, number of disabled persons in the household, number of persons currently studying in the household, the number of healthcare benefits, and the number of government assistance programs etc. Consequently, we removed the variables that did not differ significantly between cluster mean and then performed data clustering. As a results, Silhouette coefficient increased to 0.235 and 0.390 for indebted households and debt-free households respectively. We hope that the relevant government and private agencies will be able to use this research as a guide to solve the problem of debt burden among the Thai population, and to help reduce the risk of household debt in the future.

Key words: K-means, DBSCAN, household debt, data clustering, Silhouette score.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปัญหาพิเศษเล่มนี้สำเร็จลุล่วงไปได้ด้วยดีด้วยความกรุณาของ ดร.ยวุติ กลุ่มวิเศษ อาจารย์ที่ปรึกษา ที่ให้คำแนะนำ คำปรึกษา แนวคิด ให้ข้อเสนอแนะ ตลอดจนตรวจทานแก้ไขความถูกต้องและหมั่นดูแลติดตามในทุกขั้นตอน จนกระทั่งปัญหาพิเศษเล่มนี้เสร็จสมบูรณ์ ผู้วิจัยจึง ขอขอบพระคุณท่านเป็นอย่างสูงไว้ ณ ที่นี้ด้วย

ผู้วิจัยขอขอบพระคุณ ผศ.ดร.กนกวรรณ ลิ้โรจนาประภา และ ผศ.ดร.พรพิมล ชัยวุฒิศักดิ์ ผู้ซึ่งเป็น คณะกรรมการปัญหาพิเศษเล่มนี้ ที่กรุณาให้คำปรึกษาและชี้ให้เห็นถึงข้อบกพร่อง ตลอดจนแก้ไขข้อผิดพลาดเพิ่มเติม จนทำให้ปัญหาพิเศษเล่มนี้สมบูรณ์ยิ่งขึ้น

ผู้วิจัยขอขอบพระคุณคณาจารย์และบุคลากร ภาควิชาสถิติประยุกต์ทุกท่านที่ประสิทธิ์ ประสาทความรู้และประสบการณ์ ที่ได้นำมาประยุกต์ใช้ในการทำปัญหาพิเศษเล่มนี้ได้เป็นอย่างดี

ผู้วิจัยขอขอบพระคุณสำนักงานสถิติแห่งชาติ (National Statistical Office : NSO) ที่ให้ความอนุเคราะห์และการสนับสนุนด้านข้อมูล

สุดท้ายนี้ผู้วิจัยขอขอบพระคุณครอบครัวของผู้วิจัย ที่ทำการสนับสนุนและคอยให้กำลังใจ แก่ผู้วิจัย ตลอดจนเป็นแรงจูงใจที่สำคัญจนทำให้การศึกษาครั้งนี้ประสบความสำเร็จ ขอขอบคุณเพื่อน ๆ ภาควิชาสถิติประยุกต์ทุกคนที่ช่วยเป็นกำลังใจและให้ความช่วยเหลือมาโดยตลอด และผู้ที่เกี่ยวข้อง ที่ไม่ได้กล่าวนามไว้ทุกท่าน จึงขอขอบคุณมา ณ ที่นี้

พิรดาว แลซอ
วีไลร์ตัน สายสะอาด
สิริทิพย์ ทองโต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ก
บทคัดย่อภาษาอังกฤษ	ค
กิตติกรรมประกาศ	จ
สารบัญ	ฉ
สารบัญรูป	ณ
สารบัญตาราง	ญ
บทที่ 1 บทนำ	1
1.1 ความเป็นและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 แนวคิดและทฤษฎีเกี่ยวกับหนี้สิน	4
2.1.1 ความหมายของหนี้สิน	4
2.1.2 ความหมายของหนี้สินครัวเรือน	5
2.2 การทำเหมืองข้อมูล (Data Mining)	5
2.3 การค้นหาเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors)	7
2.4 การแบ่งกลุ่มข้อมูล (Data Clustering)	8
2.4.1 การจัดกลุ่มเคมีน (เคมีน Clustering)	8
2.4.2 การจัดกลุ่มเชิงพื้นที่ตามความหนาแน่นของโปรแกรมประยุกต์ที่มี การรบกวน (Density-Based Spatial Clustering of Applications with Noise; ดีบีเอสแกน)	10
2.4.3 Silhouette Analysis	12
2.5 การทดสอบความแตกต่างของค่าเฉลี่ย	13
2.5.1 การทดสอบค่าเฉลี่ย กรณีกลุ่มตัวอย่าง 1 กลุ่ม	13
2.5.2 การทดสอบค่าเฉลี่ยกรณีกลุ่มตัวอย่าง 2 กลุ่ม	13
2.6 ภาษาไพธอน (Python)	15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
2.7 แดชบอร์ด (Dashboard)	15
2.8 งานวิจัยที่เกี่ยวข้อง	16
บทที่ 3 วิธีการดำเนินงานวิจัย	20
3.1 การจัดการข้อมูล (Data Management)	20
3.1.1 การคัดเลือกข้อมูล	20
3.1.2 การทำความสะอาดข้อมูล (Data Cleaning)	23
3.1.3 การแปลงข้อมูล (Data Transformation)	23
3.1.4 จัดกลุ่มข้อมูลด้วยวิธี เคมีน และ ดีปีสแกน	24
3.1.5 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลสำหรับครัวเรือนมีภาระหนี้สิน	24
3.1.6 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลสำหรับครัวเรือนไม่มีภาระหนี้สิน	26
3.2 เครื่องมือที่ใช้ในการวิจัย	28
3.3 ขั้นตอนวิเคราะห์	28
3.3.1 ขั้นตอนที่ 1 วิเคราะห์จำนวนกลุ่มที่เหมาะสม	28
3.3.2 ขั้นตอนที่ 2 การหาค่าพารามิเตอร์ Eps ที่เหมาะสม	29
3.3.3 ขั้นตอนที่ 3 การจัดกลุ่ม (Clustering)	31
3.3.4 ขั้นตอนที่ 4 การวัดประสิทธิภาพการจัดกลุ่มด้วยวิธี Silhouette Coefficient	32
บทที่ 4 ผลการวิจัย	33
4.1 การวิเคราะห์ลักษณะทั่วไปและสัดส่วนของครัวเรือนไทยที่มีภาระหนี้และไม่มีภาระหนี้สิน	34
4.2 การวิเคราะห์แบ่งกลุ่มครัวเรือนมีภาระหนี้สิน	40
4.2.1 การแบ่งกลุ่มข้อมูลด้วยวิธี เคมีน	40
4.2.2 การแบ่งกลุ่มข้อมูลด้วยวิธี ดีปีสแกน	43
4.3 การวิเคราะห์แบ่งกลุ่มครัวเรือนไม่มีภาระหนี้สิน	44
4.3.1 การแบ่งกลุ่มข้อมูลด้วยวิธี เคมีน	44
4.3.2 การแบ่งกลุ่มข้อมูลด้วยวิธี ดีปีสแกน	47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

	หน้า
4.4 การวิเคราะห์เปรียบเทียบการแบ่งกลุ่มด้วย Silhouette Coefficient	49
4.5 การทดสอบความแตกต่างของค่าเฉลี่ย	50
4.6 แดชบอร์ด	55
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	59
5.1 สรุปผลการวิจัย	59
5.2 อภิปรายผล	60
5.3 ข้อเสนอแนะ	61
บรรณานุกรม	62
ภาคผนวก	65
ภาคผนวก ก.	66
ภาคผนวก ข.	70
ภาคผนวก ค.	72

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
1.1 แผนภาพการบริโภคเฉลี่ยของครัวเรือนที่มีสัดส่วนภาระหนี้ต่อรายได้ ณ ระดับต่างๆ	2
2.1 การทำการแบ่งกลุ่มด้วยวิธีเคมีน โดยกำหนด $K=5$	10
2.2 การกำหนด Core point สำหรับการแบ่งกลุ่มด้วยวิธีดีปัสแกน	11
2.3 การกำหนดจำนวนการแบ่งกลุ่มข้อมูลในการทำดีปัสแกน	11
2.4 วิธี Silhouette score $K=2$ ถึง 7	12
4.1 กราฟการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) เท่ากับ 2 3 และ 4 ด้วยวิธี Silhouette Score	40
4.2 กราฟหาค่า eps ที่เหมาะสมด้วยวิธี Nearest Neighbor	43
4.3 กราฟการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) เท่ากับ 2 3 และ 4 ด้วยวิธี Silhouette Score	44
4.4 กราฟหาค่า eps ที่เหมาะสมด้วยวิธี Nearest Neighbor	47
4.5 แดชบอร์ดแสดงลักษณะทั่วไปของครัวเรือนไทย	55
4.6 แดชบอร์ดแสดงลักษณะโดยรวมของครัวเรือนไทยที่มีภาระหนี้สินและไม่มีภาระหนี้สิน	56
4.7 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่มีภาระหนี้สินหน้า(1)	56
4.8 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่มีภาระหนี้สินหน้า(2)	57
4.9 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่ไม่มีภาระหนี้สินหน้า(1)	57
4.10 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่ไม่มีภาระหนี้สินหน้า(2)	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
2.1 Input สำหรับ Supervises Learning	6
2.2 Input สำหรับ Unsupervised Learning	7
2.3 ผลลัพธ์จาก Unsupervised Learning	7
3.1 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านภูมิศาสตร์และที่อยู่อาศัย	20
3.2 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านเศรษฐกิจ	21
3.3 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านสังคม	22
3.4 Python Code สำหรับการหาข้อมูลสูญหาย	23
3.5 Python Code สำหรับการแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกัน	23
3.6 ข้อมูลตัวอย่างของข้อมูลที่อยู่ในรูปมาตรฐานเดียวกัน	24
3.7 ตัวแปรด้านเศรษฐกิจที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนที่มี ภาระหนี้สิน	24
3.8 ตัวแปรด้านสังคมที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนที่มีภาระ หนี้สิน	26
3.9 ตัวแปรด้านเศรษฐกิจที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนที่ ไม่มีภาระหนี้สิน	26
3.10 ตัวแปรด้านสังคมที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนที่ไม่มี ภาระหนี้สิน	28
3.11 Python Code การวิเคราะห์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score	29
3.12 Python Code การวิเคราะห์หาค่า Eps ที่เหมาะสมด้วยวิธี Nearest Neighbors ครั้งที่ 1	29
3.13 Python Code การวิเคราะห์หาค่า Eps ที่เหมาะสมด้วยวิธี Nearest Neighbors ครั้งที่ 2	30
3.14 Python Code การแบ่งกลุ่มโดยวิธีเคมีน	31
3.15 Python Code การแบ่งกลุ่มโดยวิธีดีปีสแกน ครั้งที่ 1	31
3.16 Python Code การแบ่งกลุ่มโดยวิธีดีปีสแกน ครั้งที่ 2	32
3.17 Python Code การวัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี Silhouette Coefficient	32
4.1 ร้อยละของครัวเรือนในประเทศไทย จำแนกตามภาค เขตปกครอง	34
4.2 ร้อยละของครัวเรือนในประเทศไทย จำแนกลักษณะที่อยู่อาศัย	35

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับงานวิจัยในท้องถิ่น ผู้ใช้ควรนำเอกสารนี้ไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่	หน้า
4.3 ร้อยละของคร้วเรือนในประเทศไทย จำแนกลักษณะทางประชากรของหัวหน้าคร้วเรือน	36
4.4 ร้อยละของคร้วเรือนในประเทศไทย จำแนกลักษณะทางประชากรของคร้วเรือน	37
4.5 ร้อยละของคร้วเรือนในประเทศไทย จำแนกลักษณะทางเศรษฐกิจทั่วไปของคร้วเรือน	38
4.6 ผลการวิเคราะห์หาจำนวนกลุ่มที่เหมาะสม (K) คร้วเรือนมีภาระหนี้สินด้วยวิธี Silhouette Score	40
4.7 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมินคร้วเรือนมีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	41
4.8 ผลการวิเคราะห์หาค่า K ที่เหมาะสมคร้วเรือนไม่มีภาระหนี้สินด้วยวิธี Silhouette Score	44
4.9 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมินคร้วเรือนไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	45
4.10 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีดีปีสแกนของคร้วเรือนไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	48
4.11 การเปรียบเทียบการแบ่งกลุ่มของคร้วเรือนที่มีภาระหนี้สิน	49
4.12 การเปรียบเทียบการแบ่งกลุ่มของคร้วเรือนที่ไม่มีภาระหนี้สิน	50
4.13 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมินคร้วเรือนที่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	51
4.14 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมินคร้วเรือนที่ไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

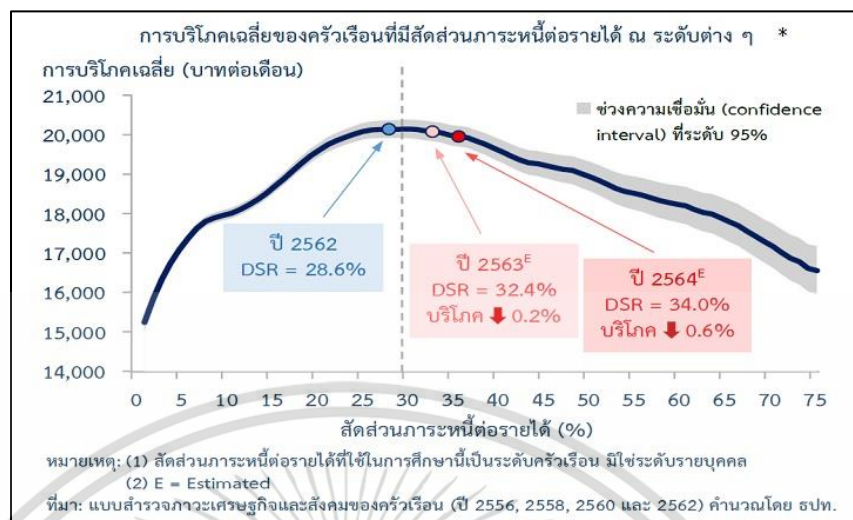
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

หนี้ครัวเรือน คือหนี้ที่ประชาชนกู้ยืมจากผู้ให้กู้ สถาบันการเงิน ผู้ให้บริการนอกระบบ หรือแม้แต่คนรู้จัก เพื่อนำเงินดังกล่าวไปใช้จ่ายตามความต้องการที่แตกต่างกันไป โดยการก่อหนี้หรือการกู้ยืมเงินนั้นช่วยให้เราสามารถใช้จ่ายได้ทั้งในชีวิตประจำวัน และยามจำเป็นเกินกว่ารายได้และเงินออมที่มีอยู่ เช่น จ่ายค่าอาหารมื้อหรูแม้เงินเดือนยังไม่ออก ซื้อรถในฝันแม้ยังไม่มีเงินออมก้อนใหญ่ หรือจ่ายค่ารักษาพยาบาลในยามฉุกเฉิน อย่างไรก็ตาม การก่อหนี้เปรียบเสมือนการนำรายได้ในอนาคตมาใช้ แม้จะทำให้เกิดการใช้จ่าย ทำให้เศรษฐกิจขยายตัวในวันนี้ แต่ในอนาคตเราจำเป็นต้องทยอยชำระหนี้คืนทำให้รายได้ที่หามาเหลือใช้ค่อยลดลง และหากครัวเรือนส่วนใหญ่ในระบบเศรษฐกิจก่อหนี้มากเกินไปก็จะส่งผลกระทบต่อเศรษฐกิจผ่านการบริโภคของครัวเรือนในอนาคตจะลดลง และความสามารถในการรองรับเหตุการณ์ไม่คาดคิดน้อยลง เช่น หากถูกเลิกจ้างหรือถูกลดค่าจ้างลง ก็อาจจะผิดนัดชำระหนี้ที่มีอยู่เดิม ซึ่งสร้างความเสี่ยงให้กับระบบสถาบันการเงินหรือผู้ให้กู้ยืม ในกรณีที่การผิดนัดชำระหนี้มีจำนวนมาก ระบบการเงินจะได้รับความเสียหายจนไม่สามารถดำเนินการได้ตามปกติและกระทบเศรษฐกิจอย่างรุนแรง (รชต และ พิรญาณ์ ,2562)

สำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน สำนักงานสถิติแห่งชาติ พบว่า (รูปที่ 1) สัดส่วนภาระหนี้ต่อรายได้ (Debt Service Ratio: DSR) โดยเฉลี่ยของครัวเรือนไทยอยู่ในจุดที่ใกล้วิกฤต (สัดส่วนภาระหนี้สินต่อรายได้ที่ระดับ 30%) มาตั้งแต่ก่อนเกิดโควิด (ปี 2562) หมายถึง ถ้าสัดส่วนนี้เพิ่มขึ้นอีกจนเกิน 30% ซึ่งเป็นจุดเปลี่ยนแปลงหรือจุดวกกลับ (turning point) โดยการก่อหนี้ในขั้นนี้จะเปลี่ยนบทบาทจากการกระตุ้นเป็นการอุดหนุนการบริโภคของครัวเรือน และเมื่อเกิดวิกฤตโควิดช่วงปี 2563 – 2564 ซึ่งส่งผลให้สัดส่วนภาระหนี้ต่อรายได้โดยเฉลี่ยจึงปรับตัวเพิ่มขึ้นเร็วเกินระดับ 30% จนทำให้ภาระหนี้กลายเป็นปัจจัยอุดหนุนการบริโภคและการฟื้นตัวของเศรษฐกิจ ซึ่งในอนาคตแม้รายได้ของครัวเรือนจะมีแนวโน้มทยอยฟื้นตัวตามกิจกรรมทางเศรษฐกิจ แต่คาดว่าภาระหนี้ที่สูงจะยังคงเป็นปัจจัยอุดหนุนเศรษฐกิจต่อไป เนื่องจากครัวเรือนจำเป็นต้องนำรายได้ที่เพิ่มขึ้นไปชำระหนี้ที่มีอยู่เดิมก่อน ทำให้ไม่สามารถใช้จ่ายได้อย่างเต็มที่ (รชต ตังนาราชชกิจ,2565)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 1.1 การบริโภคเฉลี่ยของครัวเรือนที่มีสัดส่วนภาระหนี้ต่อรายได้ ณ ระดับต่างๆ
ที่มา : แบบสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน (ปี 2556, 2558, 2560 และ 2562)
คำนวณโดยธนาคารแห่งประเทศไทย

จากปัญหาที่เกิดขึ้นในปัจจุบันภาคครัวเรือนมีสัดส่วนรายได้ต่อค่าใช้จ่ายไม่สอดคล้องกัน ซึ่งก่อให้เกิดภาระหนี้สินที่เพิ่มสูงขึ้นอย่างต่อเนื่องจนเป็นปัจจัยที่ฉุดรั้งการบริโภคและการฟื้นตัวของเศรษฐกิจ ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับลักษณะทั่วไปและสัดส่วนของภาระหนี้สินครัวเรือนในประเทศไทย เพื่อนำข้อมูลมาวิเคราะห์และจัดกลุ่มข้อมูลโดยใช้วิธีการเคมีนและดีบีสแกน (Density-Based Spatial Clustering of Applications with Noise) และนำเสนอผลการวิเคราะห์ในรูปแบบแดชบอร์ด เพื่อแสดงลักษณะภาระหนี้สินของครัวเรือนไทยตามความต้องการของผู้ใช้งาน

1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อศึกษาลักษณะทั่วไปและภาระหนี้สินครัวเรือนในประเทศไทย
- 2) เพื่อเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของคนไทย
- 3) เพื่อออกแบบแดชบอร์ดสำหรับใช้ในการนำเสนอข้อมูลภาระหนี้สินของคนไทย

1.3 ขอบเขตของงานวิจัย

งานวิจัยนี้ใช้ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ทัวราชอาณาจักร ปี พ.ศ. 2564 จากสำนักงานสถิติแห่งชาติ เป็นตัวอย่างครัวเรือนในประเทศไทย โดยมีลักษณะการศึกษาดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 1) การแบ่งกลุ่มด้วยวิธีเคมีนและดีปัสแกน (Density-Based Spatial Clustering of Applications with Noise) ใช้ในการแบ่งกลุ่ม
- 2) จัดทำแดชบอร์ดนำเสนอข้อมูลของกลุ่มครัวเรือนที่มีภาระหนี้สินและไม่มีภาระหนี้สินตามผลที่ได้จากการแบ่งกลุ่ม

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทราบวิธีการที่เหมาะสมในการแบ่งกลุ่มครัวเรือนสำหรับครัวเรือนภาระหนี้สินและไม่มีภาระหนี้สิน
- 2) ทราบตัวแปรที่เป็นลักษณะเด่นของแต่ละกลุ่มครัวเรือน และได้แดชบอร์ดที่แสดงถึงลักษณะของแต่ละกลุ่มเพื่อให้เข้าใจได้ง่ายมากยิ่งขึ้น
- 3) สามารถนำลักษณะของครัวเรือนที่แตกต่างกันในแต่ละกลุ่มใช้เป็นแนวทางประกอบการตัดสินใจในการปล่อยสินเชื่อหรือใช้เป็นแนวทางแก้ไขปัญหาหนี้สิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ศึกษาเรื่องการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของประชากรไทย โดยการประยุกต์ใช้อัลกอริทึมเคมีน, ดีปีสแกนในการแบ่งกลุ่ม และSilhouette coefficient เพื่อเปรียบเทียบความถูกต้องในการแบ่งกลุ่มครัวเรือน โดยมีวัตถุประสงค์เพื่อศึกษาจัดกลุ่มครัวเรือนตามภาระหนี้สินของคนไทย

ผู้วิจัยจึงได้ศึกษางานวิจัยภาวะเศรษฐกิจและครัวเรือนของประชากรปี พ.ศ. 2564 ของสำนักงานสถิติแห่งชาติ ทฤษฎีและงานวิจัยที่เกี่ยวข้องมาเป็นแนวทางในการศึกษาวิจัยในประเด็นต่าง ๆ ประกอบด้วยดังนี้

- 2.1 แนวคิดและทฤษฎีเกี่ยวกับหนี้สิน
- 2.2 การทำเหมืองข้อมูล (Data Mining)
- 2.3 การค้นหาเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors)
- 2.4 การแบ่งกลุ่มข้อมูล (Data Clustering)
- 2.5 การทดสอบความแตกต่างของค่าเฉลี่ย
- 2.6 ภาษา Python
- 2.7 แดชบอร์ด (Dashboard)
- 2.8 งานวิจัยที่เกี่ยวข้อง

2.1 แนวคิดและทฤษฎีเกี่ยวกับหนี้สิน

2.1.1 ความหมายของหนี้สิน

พจนานุกรมฉบับราชบัณฑิตสถานพุทธศักราช (2525) ให้นิยามคำว่า หนี้สิน คือ เงินที่ผู้หนึ่งติดค้างอยู่ซึ่งจะต้องใช้ให้แก่อีกผู้หนึ่ง

ลักษณะที่สำคัญและประเภทของหนี้สินสามารถแบ่งออกได้ดังนี้

- 1) ลักษณะที่สำคัญของหนี้สิน ดังนี้
 - เป็นภาระผูกพันในปัจจุบันที่เกิดจากเหตุการณ์ในอดีตของกิจการ เนื่องจากการประกอบธุรกิจทางการค้า เช่น การซื้อสินค้าเป็นเงินเชื่อ การให้บริการเป็นเงินเชื่อ การกู้ยืมเงินจากสถาบันการเงิน
 - เป็นหน้าที่และความรับผิดชอบที่เจ้าของกิจการต้องชดใช้ด้วยเงินสด สินทรัพย์อื่นหรือบริการตามที่ได้ตกลงกันได้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- สามารถกำหนดมูลค่าของหนี้สินและวันที่ครบกำหนดชำระหนี้ได้
 - สามารถระบุผู้รับเงินได้
- 2) ประเภทของหนี้สิน หนี้สินมีระยะเวลาในการชำระหนี้ที่แตกต่างกันขึ้นอยู่กับข้อตกลง ดังนั้น การจำแนกประเภทหนี้สินจะพิจารณาจากความแตกต่างของระยะเวลาในการชำระหนี้ ซึ่งจำแนกได้ 2 ประเภท ดังนี้
- หนี้สินหมุนเวียน (Current Liabilities) หมายถึง หนี้สินที่มีระยะเวลาครบกำหนดในการชำระคืนภายใน 12 เดือน นับจากวันที่ทำการกู้ อาจชำระด้วยเงินสดหรือสินทรัพย์หมุนเวียนอื่น
 - หนี้สินระยะยาว (Long-term Liabilities) หมายถึง หนี้สินที่มีระยะเวลาในการชำระหนี้คืนเกินกว่า 1 ปี หรือเกินกว่ารอบระยะเวลาดำเนินงานตามปกติ (รัตนภรณ์ เชาวลิตตระกูล, 2551)

2.1.2 ความหมายของหนี้สินครัวเรือน

สำนักงานสถิติแห่งชาติได้ให้คำนิยาม หนี้สินครัวเรือน หมายถึง จำนวนเงินกู้ยืมทั้งหมดของสมาชิกในครัวเรือนที่ค้างชำระ จากสถาบันการเงินและบุคคลอื่นนอกครัวเรือน และหนี้ที่เกิดจากการเช่าซื้อ การซื้อสินค้าเงินผ่อน/การซื้อสินค้าเงินเชื่อ การจำนำ การจำนอง และเงินส่งแชร์ตาย เป็นต้น

ธนาคารแห่งประเทศไทย (ธปท.) ให้คำนิยาม หนี้ครัวเรือน หมายถึง เงินให้กู้ยืมที่สถาบันการเงินให้แก่บุคคลธรรมดาที่มีถิ่นที่อยู่ในประเทศ ซึ่งบุคคลธรรมดาดำเนินการนำเงินที่กู้ยืมไปใช้เพื่อการใช้จ่ายใช้สอย หรือเพื่อประกอบธุรกิจ โดยข้อมูลหนี้ครัวเรือนจะครอบคลุมเฉพาะเงินให้กู้ยืมของสถาบันการเงินที่ธนาคารแห่งประเทศไทยไม่รวมหนี้ในระบบ

หนี้ครัวเรือน หมายถึง เงินให้กู้ยืมที่สถาบันการเงินให้แก่บุคคลธรรมดาที่มีถิ่นที่อยู่ในประเทศไทย ซึ่งบุคคลธรรมดาดำเนินการนำเงินที่กู้ยืมไปใช้เพื่อการใช้จ่ายใช้สอยต่างๆ หรือเพื่อประกอบธุรกิจ โดยข้อมูลหนี้ครัวเรือนจะครอบคลุมเฉพาะเงินให้กู้ยืมของสถาบันการเงินที่ธนาคารแห่งประเทศไทยเก็บข้อมูล ดังนั้น หนี้ครัวเรือนในที่นี้จึงไม่รวมหนี้ในระบบ ทั้งนี้สถาบันการเงิน ได้แก่ ธนาคารพาณิชย์ สถาบันการเงินเฉพาะกิจที่รับฝากเงิน สหกรณ์

2.2 การทำเหมืองข้อมูล (Data Mining)

Data Mining เป็นเทคนิคในการวิเคราะห์ข้อมูลอย่างหนึ่ง ซึ่งมาจากคำว่า เหมืองข้อมูล เป็นการค้นหาสิ่งที่มีประโยชน์จากฐานข้อมูลที่มีขนาดใหญ่ โดยเป็นกระบวนการสกัด หรือ ค้นหาสารสนเทศ เพื่อให้ได้ความรู้ หรือสารสนเทศบางมุมที่ซ่อนเร้นอยู่ในฐานข้อมูลขนาดใหญ่ ซึ่งต้องใช้ข้อมูลในอดีตเป็นจำนวนมาก เพื่อนำความรู้ที่ได้ไปช่วยในการวิเคราะห์ และประกอบการ ตัดสินใจของผู้บริหารในธุรกิจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้านการวิเคราะห์ของการสั่ง ของลูกค้า สามารถช่วยแบ่งกลุ่ม และวิเคราะห์ยอดการสั่งซื้อเพื่อนำไปผลิต และขายสินค้าได้ตรงตามกลุ่มเป้าหมายแต่ละกลุ่ม(หนึ่งททัย ชัยอากร, 2559)

ขั้นตอนการทำเหมืองข้อมูล ประกอบด้วยขั้นตอนดังนี้

1. Data Selection การเลือกแหล่งข้อมูลจากภายในและภายนอกวัตถุประสงค์
2. Data Preprocessing การเตรียมความพร้อมของข้อมูลสำหรับการใช้งาน เช่น การแก้ ความผิดพลาดหรือสุ่ม ลดขนาดข้อมูล
3. Data Transformation การแปลงข้อมูลให้อยู่ในรูปแบบที่เราต้องการ เช่น การแปลง ตัวอักษรเป็นตัวเลข
4. Data mining เอาข้อมูลมาสกัดหาสารสนเทศ มี 3 รูปแบบคือ
 - การทำนาย (Predictive Modeling) คือการสร้างแบบจำลองเพื่อการทำนาย
 - การแบ่งกลุ่มข้อมูล (Database Segmentation) คือการแบ่งกลุ่มข้อมูล
 - การศึกษากฎความสัมพันธ์ (Association Rule Discovery) คือการค้นหาค่าความสัมพันธ์ ของข้อมูลตั้งแต่สองชุดขึ้นไปภายในข้อมูลขนาดใหญ่
5. Analysis of results การวิเคราะห์ผล
6. Assimilation of knowledge กำหนดว่าใครจะใช้หรือการนำเสนอความรู้ที่ค้นพบ

โดยข้อมูลสามารถแบ่งออกเป็น 2 ประเภทคือ ซึ่ง เป็น Supervised Learning และ Unsupervised Learning โดย Supervised Learning เป็นกระบวนการแยกข้อมูลที่อิงกับข้อมูล ในอดีต ที่ได้รับการระบุกระบวนการสร้างแผนจำลองหรือเรียนรู้เกี่ยวกับคุณสมบัติของข้อมูล และ label ที่ เกี่ยวข้องเพื่อทำนาย label จากข้อมูลที่มองไม่เห็น ตัวอย่างของเทคนิค Supervised Learning มี รายละเอียดดังนี้

1. Supervised Learning คือกระบวนการเรียนรู้ที่มีการสอน โดยจะสอน Machine ว่า Inputs ของข้อมูลมีลักษณะอย่างไร แล้วจะได้ Output ของข้อมูลแบบใด อาทิเช่น เราต้องการให้ Machine เรียนรู้เรื่องเด็กนักเรียนว่าคนไหนสอบผ่านหรือสอบตก โดยหน้าตา Data Set จะมีลักษณะดังตารางที่ 2.1

ตารางที่ 2.1 Input สำหรับ Supervises Learning

ชื่อ	A	B	C	D	E	F
คะแนน	90	25	73	51	40	85
Label	Pass	Fail	Pass	Pass	Fail	?

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยในตารางที่ 2.1 Label คือการแสดง Machine ว่า Output จะมีลักษณะอย่างไร และ นำตารางนี้เข้าไปใน Machine จะทำให้สามารถเรียนรู้ได้ว่าระดับคะแนนที่เท่าไรคือผ่าน หรือ คะแนนที่เท่าไรคือตก ดังนั้น Machine สามารถคาดเดาได้ว่า นักเรียนชื่อ F มีความน่าจะเป็นที่จะ Pass เป็นผลจากข้อมูลที่ได้เรียนรู้มาโดยไม่ เป็นต้องรู้ว่าแนบเท่าไรคือผ่าน

2. Unsupervised Learning คือกระบวนการเรียนรู้แบบไม่มีการสอน โดยการเรียนรู้วิธีการจะประกอบด้วย Inputs ของข้อมูล และทำให้ Machine ทราบว่าต้องการอะไรดังตัวอย่าง เช่นการแบ่งกลุ่มในตารางที่ 2.2 ดังนี้

ตารางที่ 2.2 Input สำหรับ Unsupervised Learning

ชื่อ	A	B	C	D	E	F
คะแนน	90	25	73	51	40	85

จะเห็นได้ว่าการแบ่งเป็น 2 กลุ่ม ผลลัพธ์ที่ออกมาจะไม่สามารถทราบได้ว่า 2 กลุ่มนั้น คืออะไร คอมพิวเตอร์จะทำหน้าที่เพียงแบ่งกลุ่ม หากแบ่งเป็น 2 กลุ่ม จะได้ผลลัพธ์ดังตารางที่ 2.3

ตารางที่ 2.3 ผลลัพธ์จาก Unsupervised Learning

ชื่อ	A	B	C	D	E	F
คะแนน	90	25	73	51	40	85
Output	1	2	1	1	2	1

2.3 การค้นหาเพื่อนบ้านใกล้ที่สุด (K-nearest neighbors)

อัลกอริธึม K-nearest neighbors (KNN) เป็นการทำงานแบบการเรียนรู้ไม่มีผู้สอนใช้ในการจำแนกประเภทข้อมูล โดยหลักการการทำงาน คือ จะใช้หลักการเปรียบเทียบความคล้ายคลึงกันของข้อมูลที่สนใจกับข้อมูลอื่นที่มีความคล้ายคลึงหรืออยู่ใกล้กับข้อมูลใดมากที่สุด k ตัว จากนั้นจะทำการตัดสินใจว่าคำตอบของข้อมูลที่สนใจนั้นควรเป็นคำตอบเดียวกับข้อมูลที่อยู่ใกล้ที่สุด k ตัวนั้น ทั้งนี้ k คือความถี่ของข้อมูลที่อยู่ใกล้กับข้อมูลที่สนใจ

โดยสรุปวิธี KNN มีขั้นตอนการทำงาน ดังนี้

1. ผู้ใช้ต้องทำการกำหนดค่า k ซึ่งโดยทั่วไปจะกำหนดให้เป็นเลขคี่ เช่น 3, 5, 7 และ 9 เป็นต้น
2. นำวัตถุที่ต้องการจำแนกมาวัดหาความคล้ายคลึงหรือความต่างกับข้อมูลทั้งหมดในชุดข้อมูล โดยมาตรวัดระยะห่างที่นิยม ได้แก่ ระยะยุคลิด (Euclidean distance) ดังสมการที่ (1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{dist}(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

3. เรียงลำดับวัตถุตามความคล้ายหรือความแตกต่าง จากคล้ายคลึงมากที่สุดไปหาน้อยที่สุด หรือ จากแตกต่างน้อยที่สุดไปหามากที่สุด
4. พิจารณาคำตอบจากจำนวนคลาสคำตอบที่มีมากที่สุด ใน k ตัวที่มีความคล้ายมากที่สุด หรือมีความแตกต่างน้อยที่สุด

2.4 การแบ่งกลุ่มข้อมูล (Data Clustering)

การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งกลุ่มหน่วยข้อมูล หรือ เป็นการแบ่งคน สัตว์ สิ่งของ องค์กร ฯลฯ ออกเป็นกลุ่มย่อยอย่างน้อย 2 กลุ่ม โดยมี หลักเกณฑ์ในการแบ่งดังนี้ “ให้หน่วยที่อยู่ในกลุ่มเดียวกันมีลักษณะที่สนใจเหมือนกันหรือ คล้ายกัน แต่หน่วยที่อยู่ต่างกลุ่มกันจะมีลักษณะที่สนใจต่างกัน”

คำว่าลักษณะที่สนใจอาจจะมีหลาย ๆ ตัวแปร เช่น ถ้าสนใจความคิดเห็นทางด้าน การเมือง จะมีคำถามหลาย ๆ คำถามด้านการเมือง และจะนำคำตอบเหล่านั้นมาแบ่งกลุ่ม (กัลยา วานิชย์บัญชา, 2552)

การจัด Case (หมายถึง คน สัตว์ สิ่งของ หรือ องค์กร ฯลฯ) หรือเป็นการจัดตัวแปรออกเป็นกลุ่มย่อย ๆ ตั้งแต่ 2 กลุ่มขึ้นไป Case ที่อยู่ในกลุ่มเดียวกันจะมีลักษณะที่เหมือนกันหรือคล้ายกัน ส่วน Case ที่อยู่ต่างกลุ่มกันจะมีลักษณะที่แตกต่างกัน

ตัวแปรที่อยู่ในกลุ่มเดียวกัน จะมีความสัมพันธ์กันมากกว่าตัวแปรที่อยู่ต่างกลุ่มกัน

ตัวแปรที่อยู่ต่างกลุ่มกัน จะมีความสัมพันธ์กันน้อยหรือไม่มีความสัมพันธ์กันเลย

2.4.1 การแบ่งกลุ่มเคมีน (K-means Clustering)

เป็นเทคนิคการจำแนก Case ออกเป็นกลุ่มย่อย จะใช้เมื่อมีจำนวน Case มาก โดยจะต้องกำหนดจำนวนกลุ่มหรือจำนวน Cluster ที่ต้องการ เช่น กำหนดให้มี K กลุ่ม เคมีนจะมีการทำงานหลาย ๆ รอบ (Iteration) โดยในแต่ละรอบจะมีการรวม Cases ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่ง โดยเลือกกลุ่มที่ Case นั้นมีระยะห่างจากค่ากลางของกลุ่มน้อยที่สุด แล้วคำนวณค่ากลางของกลุ่มใหม่ จะทำเช่นนี้จนกระทั่งค่ากลางของกลุ่มไม่เปลี่ยนแปลง หรือครบจำนวนรอบที่กำหนดไว้

ชนิดของตัวแปรที่ใช้ในเทคนิคเคมีนจะต้องเป็นตัวแปรเชิงปริมาณ คือ เป็นสเกลอันตรภาค (Interval Scale) หรือสเกลอัตราส่วน(Ration Scale) โดยไม่สามารถใช้กับข้อมูลที่อยู่ในรูปความถี่ หรือ Binary เหมือนเทคนิค Hierarchical

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการวิเคราะห์ของวิธีเคมิน

การวิเคราะห์จำแนกกลุ่มด้วยเทคนิควิธีเคมิน สามารถสรุปขั้นตอนของการวิเคราะห์ที่ได้ 4 ขั้นตอน ดังนี้

ขั้นที่ 1 จัดกลุ่มข้อมูลเป็น K กลุ่ม

ขั้นที่ 2 คำนวณหาจุดกึ่งกลางกลุ่มของแต่ละกลุ่ม เช่น จุดกลางกลุ่มของกลุ่มที่ C คือ \bar{x}_c

ขั้นที่ 3 มีวิธีการพิจารณา 2 แบบ โดยจะคำนวณ

แบบที่ 1 คำนวณหาระยะห่างจากแต่ละหน่วยไปยังจุดกลางกลุ่มของทุกกลุ่ม และจะพิจารณาย้ายหน่วยไปยังกลุ่มที่มีระยะห่างต่ำสุด

แบบที่ 2 คำนวณระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่มที่หน่วยนั้นอยู่ โดยให้ ESSZ(Error Sum Square) เท่ากับระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่ม

$$ESS = \sum_{i=1}^n (x_i - \bar{x}_{c(i)}) (x_i - \bar{x}_{c(i)})$$

โดยที่ $C(i)$ หมายถึง กลุ่มของหน่วยที่ i

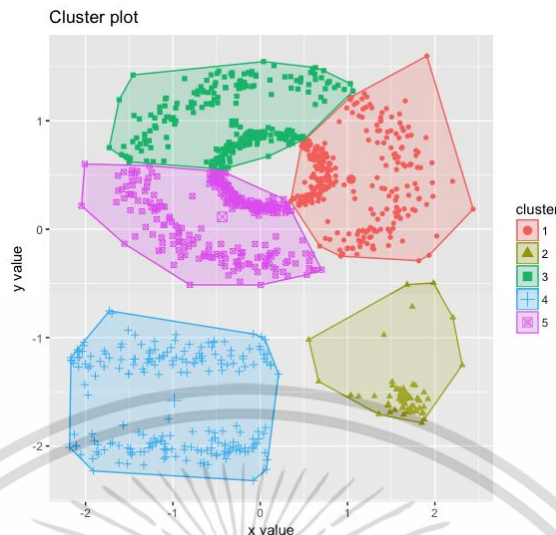
ESS = ผลบวกของระยะห่างจากแต่ละหน่วยในกลุ่มไปยังจุดกลางกลุ่มรวมทุกกลุ่ม กลุ่มใดที่มีค่า ESS ต่ำ แสดงว่าหน่วยที่อยู่ในกลุ่มนั้นมีความคล้ายคลึงกัน

ขั้นที่ 4 การพิจารณาย้ายกลุ่ม จะใช้เกณฑ์การย้ายตามค่าที่คำนวณได้ในขั้นที่ 3

แบบที่ 1 จะทำการย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ระยะห่างจากหน่วยที่ i ไปยังจุดกลางกลุ่มมีค่าต่ำสุด

แบบที่ 2 จะทำการย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ค่า ESS มีค่าต่ำสุดถ้าขั้นที่ 4 ไม่มีการย้ายกลุ่มอีกแล้ว แสดงว่ากลุ่มที่แบ่งได้นั้นเหมาะสมแล้ว แต่ถ้าในขั้นที่ 4 มีการย้ายกลุ่ม กลุ่มที่มีหน่วยย้ายเข้าหรือย้ายออกจะต้องทำการคำนวณหาจุดกลางกลุ่มใหม่นั้นคือต้องกลับไปทำขั้นที่ 2

โดยหลักการของเทคนิคเคมิน จะต้องกำหนดจำนวนกลุ่มหรือจำนวน Cluster ที่ต้องการเทคนิคเคมิน จะมีการทำงานหลาย ๆ รอบ (Iteration) โดยในแต่ละรอบจะมีการรวม Cases ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่ง โดยเลือกกลุ่มที่ Case นั้นมีระยะห่างจากค่ากลางของกลุ่มน้อยที่สุด แล้วคำนวณค่ากลางของกลุ่มใหม่ จะทำเช่นนี้จนกระทั่งค่ากลางของกลุ่มไม่เปลี่ยนแปลง หรือครบจำนวนรอบที่กำหนดไว้



รูปที่ 2.1 การทำการแบ่งกลุ่มด้วยวิธีเคมีน โดยกำหนด $K=5$
(ที่มา : Lalita Lowphansirikul, 2561)

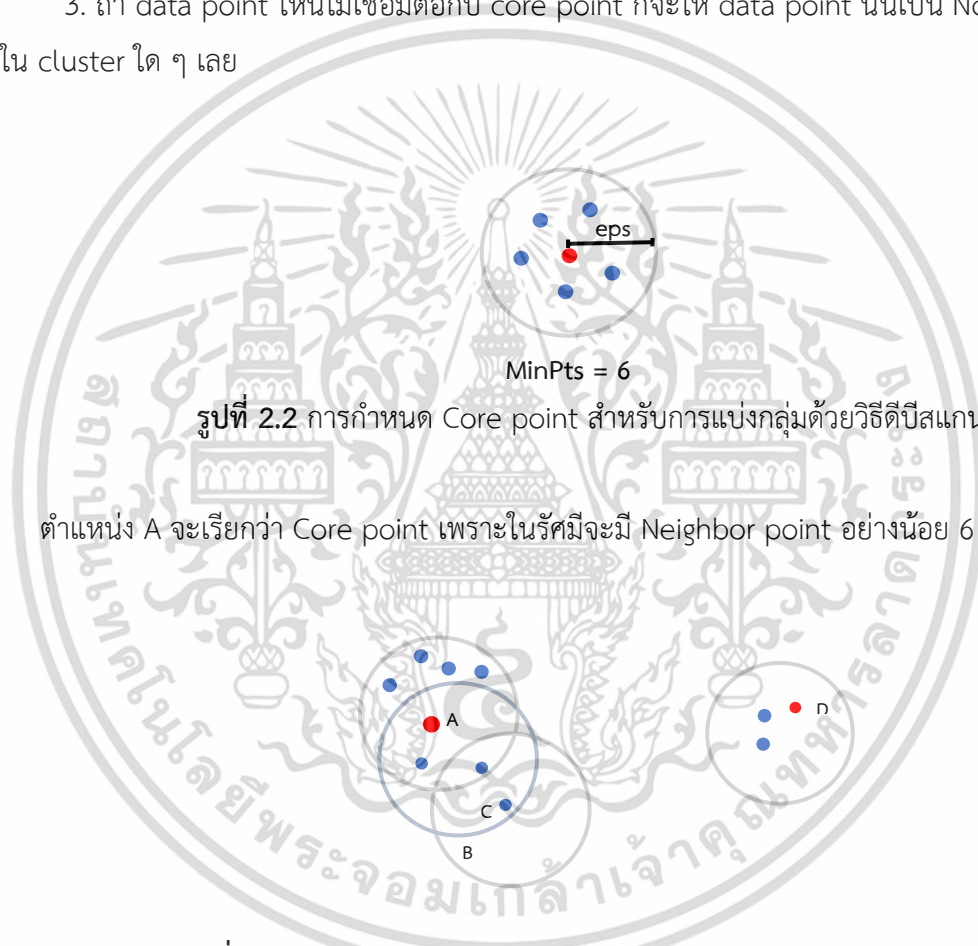
2.4.2 การแบ่งกลุ่มเชิงพื้นที่ตามความหนาแน่นของโปรแกรมประยุกต์ที่มีการรบกวน (Density-Based Spatial Clustering of Applications with Noise; DBSCAN)

ดีบีเอสแกนเป็นการแบ่งกลุ่มด้วยการพิจารณาจำนวนข้อมูลที่อยู่ในรัศมีจากจุดศูนย์กลางที่กำหนด ประกอบด้วย รัศมีของกลุ่มเพื่อนบ้านคือ การระบุรัศมีรอบข้อมูลเพื่อหาความคล้ายคลึงกันของข้อมูล และจำนวนข้อมูลน้อยที่สุด คือ การระบุจำนวนข้อมูลที่น้อยที่สุดในรัศมีเป็นจุดการเชื่อมต่อความหนาแน่นสูงสุดของการรวมกลุ่ม แนวคิดพื้นฐานดังนี้ วัตถุ p คือความหนาแน่นที่สามารถเข้าถึงได้โดยตรงจากวัตถุ q ด้วยการอ้างถึงรัศมีของกลุ่มเพื่อนบ้าน และจำนวนน้อยที่สุดในวัตถุของชุด D ถ้า p อยู่ภายในรัศมีของกลุ่มเพื่อนบ้านของ q และจะต้องมีจำนวนวัตถุที่น้อยที่สุดอยู่ในวัตถุของชุด D หากมีห่วงโซ่ของวัตถุ ดังนี้ $p_1, \dots, p_n, p_1 = q$ และ $p_n = p$ เช่นนั้นสำหรับ $1 \leq i \leq n, p_i \in D$ และ p_{i+1} เป็นความหนาแน่นที่สามารถเข้าถึงได้โดยตรงจาก p_i ด้วยการอ้างถึงรัศมีของกลุ่มเพื่อนบ้าน และจำนวนวัตถุที่น้อยที่สุด และวัตถุ p เป็นจุดการเชื่อมต่อความหนาแน่นสูงสุดถึงวัตถุ q ด้วยการอ้างถึงรัศมี และจำนวนวัตถุที่น้อยที่สุดในวัตถุของชุด D ถ้ามีวัตถุ $o \in D$ ดังนั้น p และ q คือความหนาแน่นที่สามารถเข้าถึงได้โดยตรงจาก o ด้วยการอ้างถึงรัศมีของกลุ่มเพื่อนบ้านในที่นี้เรียกว่า เอพไซลอน (Epsilon) และจำนวนจุดน้อยที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะของอัลกอริทึมดีบีเอสแกน

1. ในแต่ละ data point จะคำนวณหา neighbor point ทั้งหมดในรัศมี eps ถ้า data point ไหนมี neighbor point มากกว่าหรือเท่ากับ MinPts ให้ data point นั้นเป็น core point และสร้างเป็น cluster ใหม่
2. ในแต่ละ core point ถ้ามี neighbor ที่เชื่อมต่อกับอีก core point ได้ ให้รวมเป็น cluster เดียวกัน
3. ถ้า data point ไหนไม่เชื่อมต่อกับ core point ก็จะทำให้ data point นั้นเป็น Noise ซึ่งจะไม่อยู่ใน cluster ใด ๆ เลย



รูปที่ 2.2 การกำหนด Core point สำหรับการแบ่งกลุ่มด้วยวิธีดีบีเอสแกน

ตำแหน่ง A จะเรียกว่า Core point เพราะในรัศมีจะมี Neighbor point อย่างน้อย 6 จุด

รูปที่ 2.3 การกำหนดจำนวนการแบ่งกลุ่มข้อมูลในการทำดีบีเอสแกน

จากรูปภาพที่ 2.3 จะแสดงลักษณะในการกำหนดตำแหน่งต่าง ๆ เพื่อใช้ในการแบ่งกลุ่มของ cluster ดังนี้ ตำแหน่ง A จะเรียกว่า Core point เพราะ มี Neighbor point อย่างน้อย 6

ตำแหน่ง B จะเรียกว่า Border เพราะ B มี Neighbor point ไม่ถึง 6 แต่อยู่ในรัศมีของ core point A ตำแหน่ง C จะเรียกว่า Border เพราะ C มี Neighbor point ไม่ถึง 6 แต่อยู่ในรัศมีของ B ซึ่ง B ก็อยู่ใน

รัศมีของ Core point A ทำให้ C ถือว่าอยู่ใน cluster เดียวกันกับ A และ B

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

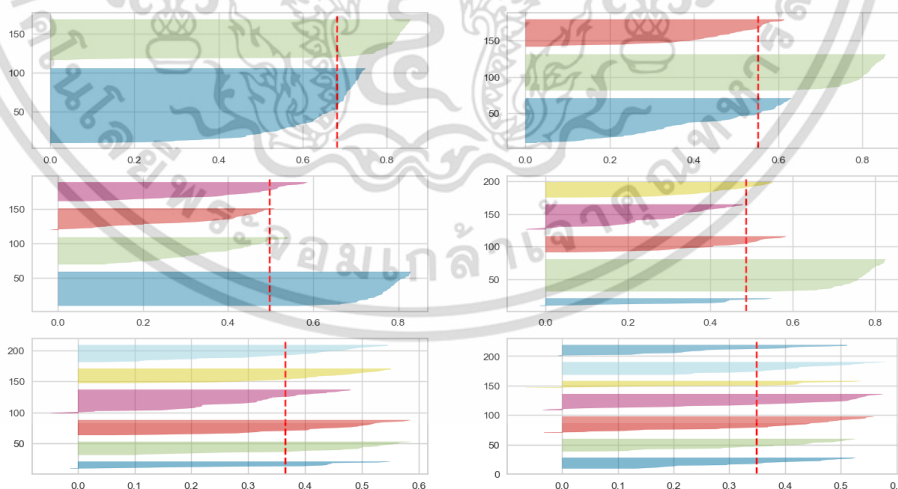
ตำแหน่ง D จะเรียกว่า Noise หรือ Outlier เพราะจุดนั้นไม่ได้อยู่ในรัศมีของ Core point ใด ๆ เลย ซึ่ง Noise นั้นจะเป็นข้อมูลที่เราต้องการตัดออกไป และไม่รวมอยู่ใน Cluster

2.4.3 Silhouette Analysis

Average silhouette score คือค่าเฉลี่ยของค่า Silhouette จากข้อมูลทั้งหมดในการวิเคราะห์กลุ่ม เป็นค่าที่บอก จำนวนกลุ่มที่เหมาะสมจากการวิเคราะห์กลุ่ม เมื่อไม่ทราบจำนวนกลุ่มที่แท้จริงหรือจำนวนกลุ่มที่ต้องการ โดย เปรียบเทียบผลการวิเคราะห์กลุ่มที่กำหนดจำนวนกลุ่มที่แตกต่างกัน (Rousseau, 1987) หากการวิเคราะห์กลุ่มที่มีจำนวนกลุ่มแตกต่างกันแบบใดมีค่า Average Silhouette Width มากกว่า แสดงว่ากลุ่มที่กำหนดนั้นมีความเหมาะสมกับข้อมูลมากกว่า กำหนดค่า Silhouette ของข้อมูลที่ i คือ

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

โดยที่ $a(i)$ แทน ระยะห่างเฉลี่ยระหว่างข้อมูลที่ i กับข้อมูลอื่น ๆ ทั้งหมดภายในกลุ่มเดียวกัน $b(i)$ แทน ระยะห่างเฉลี่ยที่น้อยที่สุดของระยะห่างเฉลี่ยระหว่างข้อมูลที่ i กับข้อมูลทั้งหมดที่อยู่กลุ่มอื่น ๆ ดังนั้น $-1 \leq S(i) \leq 1$ ซึ่งถ้า $S(i)$ มีค่าใกล้ 1 แสดงว่าข้อมูลที่ i ถูกจัดให้อยู่ในกลุ่มที่เหมาะสมแล้ว ถ้า $S(i)$ มีค่าใกล้ -1 แสดงว่าข้อมูลที่ i ควรอยู่ในกลุ่มอื่น และถ้า $S(i)$ มีค่าใกล้ 0 แสดงว่าอาจอยู่กลุ่มใดกลุ่มหนึ่งก็ได้



รูปที่ 2.4 วิธี Silhouette score K=2 ถึง 7

(ที่มา : Ajitesh Kumar, 2564)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.5 พิจารณา Silhouette score สำหรับ $n_clusters = 3$ เหมาะสมที่สุดเมื่อเทียบกับเกณฑ์การวัดทั้งสาม เนื่องจากมีผลดีต่อทั้งสามเกณฑ์การวัด คือ แต่ละคลัสเตอร์มีคะแนนที่สูงกว่าคะแนน Silhouette เฉลี่ย, มีความผันผวนน้อยในขนาดของโครงร่างภาพเงา และ ความหนาของภาพเงามีความสม่ำเสมอทั้ง 3 คลัสเตอร์

2.5 การทดสอบความแตกต่างของค่าเฉลี่ย

2.5.1 การทดสอบค่าเฉลี่ย กรณีกลุ่มตัวอย่าง 1 กลุ่ม

ในการทดสอบค่าเฉลี่ยกรณีกลุ่มตัวอย่างมี 1 กลุ่มจะเป็นการทดสอบความแตกต่างของค่าเฉลี่ยกับค่าคงที่ค่าหนึ่งที่ผู้วิจัยสนใจที่ต้องการเปรียบเทียบ ซึ่งค่าคงที่นี้อาจได้มาจากการกำหนดขึ้นหรือการทบทวนวรรณกรรมที่เกี่ยวข้องในเรื่องนั้นๆ ซึ่งการใช้สถิติทดสอบ t - test ทดสอบค่าเฉลี่ยกรณีกลุ่มตัวอย่าง 1 กลุ่ม มีสูตรในการคำนวณ ดังนี้

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \rightarrow t = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} ; df = n - 1$$

สมมติฐานทางสถิติ

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

2.5.2 การทดสอบค่าเฉลี่ยกรณีกลุ่มตัวอย่าง 2 กลุ่ม

ในการทดสอบค่าเฉลี่ยกรณีกลุ่มตัวอย่าง 2 กลุ่มนั้นจะพิจารณาว่า กลุ่มตัวอย่างทั้งสองกลุ่มเป็นอิสระจากการหรือไม่ เพื่อเลือกใช้สูตรของสถิติทดสอบให้ถูกต้อง นอกจากนี้ยังพิจารณาอีกว่าความแปรปรวนของประชากรของกลุ่มตัวอย่างเท่ากันหรือไม่ ซึ่งในการใช้สถิติ t-test ทดสอบกรณีกลุ่มตัวอย่าง 2 กลุ่ม ที่เป็นอิสระต่อกันนั้นมีสูตรที่ใช้ทดสอบอยู่ 2 สูตรด้วยกัน กล่าวคือสูตรที่ใช้ในกรณีความแปรปรวนของประชากร 2 กลุ่มมีค่าเท่ากัน ($\sigma_1^2 = \sigma_2^2$) หรือในกรณีกลุ่มตัวอย่างแต่ละกลุ่มมีจำนวนเท่ากัน (t-test แบบ Pooled variance) และสูตรที่ใช้ในกรณีความแปรปรวนของประชากร 2 กลุ่ม มีค่าไม่เท่ากัน ($\sigma_1^2 \neq \sigma_2^2$) (t-test แบบ Separated variance) ดังนั้นเมื่อผู้วิจัยจะใช้ t-test กรณีดังกล่าวจะต้องทำการทดสอบก่อนว่า ความแปรปรวนของประชากรแต่ละกลุ่มมีค่าเท่ากันหรือไม่โดยใช้ F-test เพื่อจะได้เลือกใช้สูตรของ t-test ได้อย่างถูกต้องและเหมาะสมต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมติฐานทางสถิติ

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

กรณีกลุ่มตัวอย่างเป็นอิสระต่อกัน

- เมื่อสุ่มตัวอย่างขนาด n_1 และ n_2 มาโดยอิสระจากกัน มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ μ_1 และ μ_2 ความแปรปรวนเท่ากับ σ_1^2 และ σ_2^2 ซึ่งไม่ทราบค่า แต่ทราบว่า $\sigma_1^2 = \sigma_2^2$ โดย n_1 และ n_2 น้อยกว่า 30 ใช้สูตร t - test (t-test แบบ Pooled variance)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{sp^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

เมื่อ

$$Sp^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- เมื่อสุ่มตัวอย่างขนาด n_1 และ n_2 มาโดยอิสระจากกัน มีการแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ μ_1 และ μ_2 ความแปรปรวนเท่ากับ σ_1^2 และ σ_2^2 ซึ่งไม่ทราบค่าแต่ทราบว่า $\sigma_1^2 \neq \sigma_2^2$ โดย n_1 และ n_2 น้อยกว่า 30 ใช้สูตร t - test (t-test แบบ Separated variance)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}; df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\left(\frac{s_1^2}{n_1} \right) + \left(\frac{s_2^2}{n_2} \right)}$$

กรณีกลุ่มตัวอย่างมีความสัมพันธ์กัน

การเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย กลุ่มที่มีความสัมพันธ์กัน (ไม่อิสระจากกัน) เช่น ในกรณีที่ทำการศึกษาที่มีลักษณะเป็นคู่กัน เช่น ฝาแฝด หรือคนกลุ่มเดียวแต่มีการทดสอบสองครั้ง เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทดสอบก่อนการทดลองและทดสอบหลังการทดลอง โดยใช้ค่าแจกแจง t – test แบบ Dependent Samples ดังนี้

$$t = \frac{\sum D}{\sqrt{\frac{N \sum D^2 - (\sum D)^2}{N-1}}} ; df = n - 1$$

2.6 ภาษาไพธอน (Python)

ภาษาไพธอน (Python) เป็นภาษาการเขียนโปรแกรมระดับสูง ที่นำข้อดีของภาษาต่างๆ มารวมไว้ด้วยกันถูกออกแบบมาให้เรียนรู้ได้ง่าย และมีไวยากรณ์ที่ช่วยให้เขียนโค้ดสั้นกว่าภาษาอื่นๆ มีความสามารถใช้นิพจน์ข้อมูลแบบไดนามิก จัดการหน่วยความจำอัตโนมัติ สนับสนุนกระบวนทัศน์การเขียนโปรแกรม (Programming paradigms) ประกอบด้วย การเขียนโปรแกรมเชิงวัตถุ (OOP : Object Oriented Programming) การเขียนโปรแกรมเชิงคำสั่ง (Imperative Programming) การเขียนโปรแกรมเชิงฟังก์ชัน (Functional) และการเขียนโปรแกรมเชิงกระบวนการ มีลักษณะเป็นภาษาสคริปต์ที่ทำงานร่วมกับภาษาอื่นได้ มีไลบรารีมาตรฐานมากมาย และใช้อินเตอร์พรีเตอร์แปลภาษาโปรแกรมให้ทำงานบนระบบปฏิบัติการได้หลากหลาย ทั้งบน Windows, MAC, Linux และ Unix นอกจากนี้ยังเป็นโปรแกรมแบบ Open source ที่นำใช้ได้ฟรี เหมาะสำหรับโปรแกรมทั้งขนาดเล็กและขนาดใหญ่ เช่น การสร้างเกม เพรมเวิร์กพัฒนาเว็บ โปรแกรมที่ใช้กราฟิกติดต่อกับผู้ใช้งาน (GUI) งานคำนวณทางวิทยาศาสตร์ และสถิติ งานพัฒนาซอฟต์แวร์ และซอฟต์แวร์ควบคุมระบบ เป็นต้น

2.7 แดชบอร์ด (Dashboard)

Dashboard คือหน้ากระดานที่ใช้ในการสรุปข้อมูลแบบ Executive ในมุมมองต่าง ๆ เพื่อให้สามารถดูได้ง่าย ๆ ใช้เวลาในการตีความสั้น ๆ และสามารถตอบโจทย์ในทางธุรกิจได้ ใช้ในการติดตามเรื่องที่น่าสนใจ เพื่อเห็นการเปลี่ยนแปลงของข้อมูลตลอดเวลา โดย Dashboard มีทั้งหมด 4 ประเภทดังนี้

1. Strategic Dashboards เป็น Dashboard เชิงกลยุทธ์
2. Analytical Dashboards เป็น Dashboard เชิงวิเคราะห์
3. Operational Dashboards เป็น Dashboard ที่ใช้ตรวจสอบการทำงานอย่างต่อเนื่อง
4. Informational Dashboards เป็น Dashboard ที่แสดงข้อมูล ตัวเลขที่เป็นข้อเท็จจริงหรือสถิติ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หลักการงานแบบของ Dashboard จะมีอยู่ 4 ขั้นตอนหลัก ดังนี้

1. นำเข้าข้อมูลต่าง ๆ ออกมาจาก Database เช่น SQL / SAP / Google Analytics / Facebook/ รายงานประจำเดือนหรือช่องทางอื่นๆ
2. เมื่อได้ข้อมูลก็จะทำการใช้โปรแกรม Business Intelligence (BI) เช่น Power BI, Tableau เป็นต้น ในการ Transform และ Join ข้อมูลจากหลายๆ ที่โดยการทำให้ Data Model
3. จากข้อมูลที่ได้ ก็สามารถนำเสนอในรูปแบบของกราฟ เช่น เปรียบเทียบยอดขายรายช่องทาง เปรียบเทียบยอดขายกับเป้าที่วางไว้ การเปลี่ยนแปลงของยอดขายจากเดือนที่แล้ว หรือปีที่แล้ว นอกจากกราฟต่างๆ แล้ว ยังสามารถทำ conditional formatting ของตารางได้ เช่น ตารางยอดขายรายพนักงานขาย ให้เทียบกับเป้า แล้วใครที่ขายได้ถึงเป้าให้แสดงสีเขียว ใครขายไม่ถึงเป้าให้แสดงสีแดง
4. เมื่อถึงเดือนใหม่ แทนที่จะต้องทำ Report ก็สามารถใช้ Dashboards นี้ ขึ้น Present Report ให้กับผู้บริหารหรือผู้ที่เกี่ยวข้อง รวมถึงการประหยัดเวลา จะช่วยให้ได้มุมมองใหม่ ๆ และการวิเคราะห์ที่ซับซ้อนได้มากยิ่งขึ้น

2.8 งานวิจัยที่เกี่ยวข้อง

วนิดา แผลงรักษา และ นิเวศ จิระวิชิตชัย (2565) ได้ทำการศึกษาการบริหารลูกค้า โดยมีวัตถุประสงค์เพื่อนำเสนอการแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีนสำหรับการบริหารลูกค้าสัมพันธ์โดยทดสอบกับระบบฐานข้อมูลเกี่ยวกับลักษณะลูกค้าของผู้จัดจำหน่ายสินค้าออนไลน์แห่งหนึ่ง โดยพิจารณาจากคำสั่งซื้อสินค้าของลูกค้าทั้งหมดจำนวน 1,000 คน เพื่อนำมาวิเคราะห์การแบ่งกลุ่มคลัสเตอร์ด้วยเคมีน อาทิ ข้อมูลสถานภาพ ข้อมูลเพศ ข้อมูลรายได้ต่อเดือน จำนวนบุตร การสำเร็จการศึกษา อาชีพ ระยะทางระหว่างที่พัก-ที่ทำงาน ภูมิลำเนาและจำนวนรถที่เป็นเจ้าของนำวิเคราะห์ลักษณะทางประชากรศาสตร์ ผลจากการทดลองโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน พบว่าสามารถแบ่งกลุ่มลูกค้าที่มีคุณลักษณะเหมือนกันไว้ในกลุ่มเดียวกันได้อย่างมีประสิทธิภาพ และทราบว่าลูกค้าสามารถสังเคราะห์และจำแนกออกมาได้ตามลักษณะอิทธิพลต่อทางพฤติกรรมของกลุ่มลูกค้าว่ามีประสิทธิภาพดีในการใช้เทคนิคการทำคลัสเตอร์แบบเคมีนในระดับ 7 กลุ่ม ตามลักษณะคุณสมบัติเด่น ๆ ในการกำหนดแผนการส่งเสริมการตลาดที่เฉพาะเจาะจงให้เกิดผลสูงสุดเพื่อกระตุ้นให้เกิดการซื้อสินค้า และชักจูงใจให้ผู้บริโภคมีแรงจูงใจในการสั่งซื้อสินค้าและบริการ สำหรับการบริหารลูกค้าสัมพันธ์ ซึ่งจะทำให้สามารถกำหนดกลยุทธ์ทางการตลาดเพื่อจับกลุ่มเป้าหมายได้ถูกต้องแม่นยำและมีประสิทธิภาพมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

มาริสสา สกุลวัฒนา (2564) ได้ทำการศึกษาพฤติกรรมการก่อกวนที่ก่อให้เกิดพฤติกรรมการใช้จ่าย และเพื่อนำเสนอแนวทางการแก้ปัญหาการก่อกวนของกลุ่มพนักงานวัยเริ่มทำงานของบริษัทเอกชน เลือกการวิเคราะห์ข้อมูลออกเป็น 2 ส่วน ได้แก่ การวิเคราะห์ข้อมูลโดยใช้สถิติเชิงอนุมาน (Inferential Statistics Analysis) และสถิติเชิงวิเคราะห์(analytical statistics) โดยใช้การทดสอบไคสแควร์(Chisquare) พบว่าตัวอย่างส่วนใหญ่ มีค่าใช้จ่ายในการซื้ออาหาร การบริโภคมากที่สุด มีค่าใช้จ่ายในการผ่อนซื้ออุปกรณ์เครื่องใช้ไฟฟ้า เครื่องอำนวยความสะดวก ค่าใช้จ่ายในการซื้อเสื้อผ้า เครื่องแต่งกาย ค่าใช้จ่ายในการเดินทางในระดับปานกลาง โดยที่กลุ่มตัวอย่างส่วนใหญ่ไม่มีเงินสำรองสำหรับเหตุการณ์ฉุกเฉิน มีการออมเงินฝากในธนาคาร และการลงทุนค่อนข้างน้อย และกลุ่มตัวอย่างส่วนใหญ่มีหนี้ในระบบมากที่สุด โดยเฉพาะแหล่งกู้จากสถาบันการเงิน และมีการชำระหนี้ทั้งเต็มจำนวน และชำระหนี้ขั้นต่ำ

รุจิรา จุลภักดิ์, จิรภา เพชรพัฒนานันท์ และ ศิริเรือง พัฒน์ช่วย (2564) ได้ทำการศึกษาหนี้สินของเกษตรกร โดยมีวัตถุประสงค์เพื่อออกแบบและพัฒนาโครงข่ายประสาทเทียมสำหรับพยากรณ์หนี้สินของเกษตรกร ผู้วิจัยได้นำทฤษฎีการเรียนรู้ของนิเวรอลเน็ตเวิร์คมาประยุกต์ใช้กับข้อมูลกลุ่มตัวอย่างที่ถูกเก็บรวบรวมจากเกษตรกรประกอบด้วยข้อมูลรายรับรายจ่ายของเกษตรกร จำนวน 100 ครัวเรือน ระยะเวลาในการจัดเก็บ 1 ปี รวมข้อมูลทั้งสิ้น 36,500 ชุดข้อมูล มี 17 ปัจจัย ได้แก่ ข้อมูลเงินได้สุทธิ เงินเก็บก่อนใช้ อื่นๆ ค่าอาหาร ค่าที่พัก ค่าน้ำ ค่าไฟ ค่าเครื่องนุ่งห่ม ค่ายา ค่ายานพาหนะ ค่าลอตเตอรี่ ค่าการพนัน ค่าสุรา ค่าบุหรี่ เงินต้นหนี้สิน ดอกเบี้ย ค่าท่องเที่ยว ค่าทำบุญ เพื่อใช้ในการทดสอบในงานวิจัยครั้งนี้ ผลการวิจัยพบว่า จำนวนโครงข่ายประสาทเทียมที่อยู่ในโครงสร้างถูกจัดเรียงในลักษณะการเรียนรู้แบบแพร่กระจายไปข้างหน้า (Feed-Forward Neural Network) โดยมีค่าเฉลี่ยของความถูกต้องเท่ากับ 99.1537% ,ค่าความผิดพลาดจากการจำแนกกลุ่ม เท่ากับ 0.8463% ซึ่งผลลัพธ์อยู่ในเกณฑ์ระดับมาก และทำให้เกิดองค์ความรู้ใหม่จากงานวิจัย สามารถนำไปเผยแพร่ไปสู่ประชากรเพื่อนำไปประยุกต์ใช้งาน การประเมินความพึงพอใจของโมเดลที่ใช้ในการวิเคราะห์พยากรณ์หนี้สินจากบัญชีครัวเรือนจากกลุ่มเกษตรกร โดยมีความคิดเห็นโดยรวมอยู่ในระดับมากที่สุด (\bar{X} = 4.54, S.D.= 0.35) เมื่อพิจารณารายข้อพบว่าประเด็นผลการพยากรณ์สามารถนำไปประกอบการตัดสินใจ มีค่าเท่ากับ (\bar{X} = 4.68, S.D.= 0.56) อยู่ในระดับมากที่สุด

อารีรัตน์ ลุนลลาด และ อมรรัตน์ อภินันท์มหกุล (2564) ได้ทำการศึกษาปัญหาหนี้สินประเภทต่าง ๆ ของครัวเรือนทั่วประเทศ อาศัยข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมครัวเรือน (Socio Economic Survey : SES) ของสำนักงานสถิติแห่งชาติ ปี พ.ศ. 2558 มุ่งศึกษาประเด็นสำคัญในเรื่องหนี้สินด้านการเกษตร เพื่อวิเคราะห์หาปัจจัยสำคัญอันเป็นสาเหตุของปัญหาหนี้สินของครัวเรือนที่ประกอบอาชีพด้านการเกษตร โดยเฉพาะอย่างยิ่งครัวเรือนที่ทำนาและหาแนวทางการแก้ไขปัญหาหนี้สินของเกษตรกรอย่างยั่งยืน ใช้วิธีการประมาณค่าด้วยแบบจำลองถดถอยโทบิท (Tobit Regression) ที่เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถแก้ปัญหาข้อจำกัดของข้อมูลแบบ Censoring Data ที่ใช้ในการวิจัยได้ ผลการศึกษาพบว่า ตัวแปรสำคัญต่อภาระหนี้สินครัวเรือนคือการมีสมาชิกที่จบปริญญาตรี รายได้ต่อหัวของครัวเรือน มูลค่าสินทรัพย์ทั้งหมด เงินรับโอนในรอบปีและค่าใช้จ่ายในครัวเรือน ประเด็นที่น่าสนใจที่พบจากการศึกษาในครั้งนี้ คือ ครัวเรือนที่มีที่ดินทำการเกษตรมากส่งผลให้มีภาระหนี้สินด้านการเกษตรเพิ่มขึ้น สะท้อนว่าอาชีพด้านการเกษตรมีต้นทุนที่สูงและยังทำการเกษตรมากภาระค่าใช้จ่ายจะมากขึ้นเช่นกัน

อาเดรียโน โด. อมาเรอา, วิลลาวัลด์ ฟิงค์ และ เต็ง จิงหยู (2563) ได้ทำการศึกษาอธิบายถึงผลกระทบจากการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ที่มีผลต่อหนี้ครัวเรือนไทย โดยมีวัตถุประสงค์เพื่ออธิบายผลกระทบจากเหตุการณ์การระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ที่มีผลกระทบต่อหนี้ครัวเรือนไทย โดยข้อมูลที่ใช้ในการศึกษานี้มาจากสำนักดัชนีเศรษฐกิจการค้า กระทรวงพาณิชย์ สำนักงานสภาพัฒนาการเศรษฐกิจและสังคมแห่งชาติ และธนาคารแห่งประเทศไทย ในช่วงระหว่างไตรมาสแรกของปี 2556 ถึงไตรมาสที่สองของปี 2564 โดยการวิเคราะห์ข้อมูลเชิงปริมาณแบบการวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square; OLS) ผลการศึกษาพบว่าในช่วงการแพร่ระบาด ตั้งแต่ไตรมาสที่ 1 ปี 2563 ถึงไตรมาสที่ 2 ปี 2564 ยอดหนี้ครัวเรือนรวมสูงขึ้น 5.2% เมื่อเทียบกับช่วงก่อนการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ครัวเรือนไทยมีความเสี่ยงต่อการผิดนัดชำระหนี้เพิ่มขึ้น 2.74% ในช่วงการระบาดใหญ่และวิกฤตโรคระบาดทำให้สินเชื่อภาคครัวเรือนชะลอตัว การเติบโตของสินเชื่อลดลง 0.79% ต่อไตรมาส เมื่อเทียบกับระดับการเติบโตต่อไตรมาสก่อนการระบาด ผลกระทบที่รวมกันเหล่านี้อาจนำไปสู่ความไม่มั่นคงทางการเงินและการบริโภคที่ลดลง ทำให้การฟื้นตัวจากภาวะถดถอยภายหลังการระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ล่าช้าออกไป

สำนักงานสถิติ (2562) ได้ทำการศึกษาหนี้สินของครัวเรือนเกษตร ปี พ.ศ. 2562 โดยมีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีอิทธิพลต่อการเป็นหนี้ของครัวเรือนเกษตร โดยใช้สถิติวิเคราะห์ การถดถอยโลจิสติกเชิงพหุแบบทวิภาค พบว่ามี 9 ตัวแปร จากทั้งหมด 13 ตัวแปร ที่มีผลต่อการเป็นหนี้ของครัวเรือนเกษตร ได้แก่ ภาค สถานภาพสมรสของหัวหน้าครัวเรือน อายุของหัวหน้า ครัวเรือน ระดับการศึกษาสูงสุดของหัวหน้าครัวเรือน จำนวนสมาชิกในครัวเรือน ค่าใช้จ่ายทั้งสิ้น เฉลี่ยต่อเดือน มูลค่าทรัพย์สินเฉลี่ย จำนวนพื้นที่ทำการเกษตร และการเป็นเจ้าของที่ดิน โดยมีรายละเอียดดังนี้ ครัวเรือนเกษตรที่อยู่ในภาคตะวันออกเฉียงเหนือ มีโอกาสเป็นหนี้มากกว่าครัวเรือนเกษตรในภาคอื่น ๆ หัวหน้าครัวเรือนเกษตรที่มีสถานภาพเป็นสมรส มีโอกาสเป็นหนี้มากกว่าสถานภาพอื่น ๆ หัวหน้าครัวเรือนเกษตรที่จบประถมศึกษาหรือต่ำกว่ามีโอกาสเป็นหนี้น้อยกว่าหัวหน้า ครัวเรือนเกษตรที่จบมัธยมศึกษาขึ้นไป ถ้าหัวหน้าครัวเรือนเกษตรมีอายุเพิ่มขึ้น 1 ปี มีโอกาสเป็นหนี้ลดลง ถ้าครัวเรือนเกษตรมีจำนวนสมาชิกเพิ่มขึ้น 1 คน มีโอกาสที่จะเป็นหนี้เพิ่มขึ้น ถ้าครัวเรือนเกษตรมีจำนวนพื้นที่ทำการเกษตรเพิ่มขึ้น 1 ไร่ มี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โอกาสเป็นหนี้เพิ่มขึ้น ครวเรือนเกษตรที่ไม่มีที่ดินเป็นของตนเอง มีโอกาสเป็นหนี้มากกว่าครวเรือนเกษตรที่มีที่ดินเป็นของตนเอง ส่วนครวเรือนเกษตรที่ไม่ใช้ที่ดินในการทำเกษตร มีโอกาสเป็นหนี้น้อยกว่าครวเรือนเกษตรที่มีที่ดินเป็นของตนเอง ค่าใช้จ่ายทั้งสิ้นเฉลี่ยต่อเดือนและมูลค่าทรัพย์สินเฉลี่ย แม้ว่าจะมีความสัมพันธ์กับการเป็นหนี้ของครวเรือนเกษตรอย่างมีนัยสำคัญทางสถิติ

ชฎารัตน์ พิพัฒน์นันท และ สิริธร เจริญรัตน์ (2559) ได้ทำการศึกษาข้อมูลทางเศรษฐกิจและสังคม โดยงานวิจัยนี้เป็นการศึกษาและวิเคราะห์ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคม ของครวเรือนประจำปี พ.ศ. 2552 โดยนำความรู้ทางด้านการทำเหมืองข้อมูลมาใช้เพื่อจัดกลุ่มข้อมูลตามการใช้จ่ายของประชากรโดยใช้อัลกอริทึมเคมีนสำหรับหาคลุ่มของข้อมูลที่เหมาะสมจากค่า DB Index และ SD Validity Index พบว่าจำนวนคลุ่มที่เหมาะสม คือ 3 คลุ่ม จากนั้นใช้กฎความสัมพันธ์ในการหาความสัมพันธ์ของข้อมูลในแต่ละคลุ่ม เพื่อนำไปหาตัวแปรที่มีความสัมพันธ์กัน ผลการวิจัย พบว่า กฎความสัมพันธ์ที่ได้ในแต่ละคลุ่มมีลักษณะคล้ายคลึงกัน เช่น รายได้เฉลี่ยต่อเดือนของครวเรือนมีความสัมพันธ์กับค่าใช้จ่ายเฉลี่ยต่อเดือนของครวเรือน ขนาดของครวเรือนมีความสัมพันธ์กับจำนวนผู้หารายได้ และค่าใช้จ่ายยาสูบต่อเดือนของครวเรือนมีความสัมพันธ์กับจำนวนสมาชิกที่มีสิทธิในการเบิกค่ารักษาพยาบาล เป็นต้น

BENMOUIZA Khalil และ CHEKNANE Ali (2559) ได้ทำการศึกษานุกรมเวลาการแพร่รังสีแสงอาทิตย์ โดยมีวัตถุประสงค์เพื่อศึกษาพฤติกรรมพลวัตของสุริยะ โดยใช้ DBSCAN (Density-Based Spatial Clustering of Applications with Noise) ใช้เพื่อจัดกลุ่มนุกรมเวลาการแพร่รังสีแสงอาทิตย์ และตรวจจับข้อมูลรบกวนนอกจากนี้วิธีการที่เสนอจะถูกเปรียบเทียบกับ 2 เทคนิค (unsupervised clustering) คือ K-means และ Fuzzy c-means สำหรับการวิเคราะห์นุกรมเวลาการแพร่รังสีดวงอาทิตย์รายชั่วโมงที่วัดได้ จากผลการวิเคราะห์พบว่าอัลกอริทึม DBSCAN มีการจัดการข้อมูลรบกวนประมาณ 10% ตรงกันข้ามกับ K-means และ FCM ที่ไม่มีการจัดการข้อมูลรบกวนที่ชัดเจนเนื่องจากทุกจุดถูกกำหนดให้กับคลัสเตอร์ นอกจากนี้อัลกอริทึม K-means ยังใช้เวลาที่ผ่านไปน้อยกว่า (0.45 วินาที) มากกว่าอัลกอริทึม FCM ซึ่งใช้เวลา 1.22 วินาที และ DBSCAN ที่ใช้เวลา 1.58 วินาทีเนื่องจากการคำนวณมากในแต่ละอัลกอริทึม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินงานวิจัย

ปัญหาพิเศษเล่มนี้ มีวัตถุประสงค์เพื่อศึกษาภาระหนี้สินของประชากรไทยโดยใช้การแบ่งกลุ่มแบบเคมีนและดีปัสแกน เพื่อศึกษาถึงปัจจัยที่มีผลต่อภาระหนี้สินครัวเรือน โดยนำข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2564 ซึ่งเป็นข้อมูลทุติยภูมิ รวบรวมโดยสำนักงานสถิติแห่งชาติ จากแบบสำรวจครัวเรือนส่วนบุคคล ทั้งในเขตและนอกเขตเทศบาลทั่วประเทศ จำนวน 46,840 ครัวเรือน ในเดือนมกราคม - ธันวาคม พ.ศ. 2564

ข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ปี 2564 มี 46,840 ตัวอย่าง 888 ตัวแปร ซึ่งไม่สามารถนำไปวิเคราะห์ได้ จึงต้องมีการตัดตัวแปรบางตัวแปรออกและทำการเปลี่ยนชื่อตัวแปร

3.1 การจัดการข้อมูล (Data Management)

3.1.1 การคัดเลือกข้อมูล

เป็นขั้นตอนการเลือกข้อมูลที่จำเป็นจากฐานข้อมูลจากสำนักงานสถิติแห่งชาติเพื่อใช้ในการวิเคราะห์เลือกเฉพาะข้อมูลที่เกี่ยวข้องกับวัตถุประสงค์ของการวิเคราะห์ โดยทำการเลือกข้อมูลคำถามที่มีผู้ตอบแบบสอบถามตอบครบถ้วน พร้อมทั้งคัดเลือกข้อมูลคำถามที่มีความเกี่ยวข้องกับงานวิจัยนี้โดยทำการตัดข้อมูลคำถาม เนื่องจากเป็นคำถามที่เจาะจงถามเฉพาะคนใดคนหนึ่งในครัวเรือนและข้อมูลที่ได้ทำการเก็บรวบรวมไว้เป็นข้อมูลเชิงคุณภาพ เช่น เพศ(สมาชิกในครัวเรือน) สถานภาพการทำงาน ลักษณะที่อยู่อาศัย เป็นต้น นอกจากนั้นจึงได้กำหนดชื่อตัวแปรใหม่จะเหลือข้อมูลดังต่อไปนี้

ตารางที่ 3.1 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านภูมิศาสตร์และที่อยู่อาศัย

ชื่อตัวแปร	คำอธิบาย
REG	ภาค
CWT	จังหวัด
AREA	เขตเทศบาล
Habitation	ประเภทที่อยู่อาศัย
House_Occupation	สถานภาพการครองครองที่พัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.2 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านเศรษฐกิจ

ชื่อตัวแปร	คำอธิบาย
AE_Debt	เป็นหนี้หรือไม่เป็นหนี้
Value_Land	มูลค่า (บ้าน ที่ดิน และสิ่งปลูกสร้าง)ที่ครัวเรือนเป็นเจ้าของ
Value_Vehicle	มูลค่า (ยานพาหนะทุกประเภท) ที่ครัวเรือนเป็นเจ้าของ (ทั้งใช้ในครัวเรือน และใช้ทำธุรกิจ)
Value_HouseAsset	มูลค่าสินทรัพย์ทางการเงินที่ครัวเรือนเก็บออม
Issue_Cost	ใน 12 เดือนที่แล้ว ครัวเรือนมีปัญหาในการจ่ายค่าเช่าบ้าน ค่าน้ำ ค่าไฟ ค่าเล่าเรียนหรือไม่
Borrow_Farm_Business	ในกรณีที่ท่านต้องการกู้หรือยืมเงินเพื่อลงทุนทำธุรกิจหรือทำการเกษตร ท่านสามารถกู้หรือยืมเงินได้หรือไม่
Borrow_Emergency	ในกรณีที่ท่านต้องการกู้หรือยืมเงินเพื่อใช้จ่ายฉุกเฉิน ท่านสามารถกู้หรือยืมได้หรือไม่
Source_Loan	แหล่งกู้ยืมหนี้สิน
Habitation_FM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้ในระบบ)
Education_FM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้ในระบบ)
Consumer_FM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้ในระบบ)
Business_FM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้ในระบบ)
Farm_FM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้ในระบบ)
Other_FM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย ฯลฯ (หนี้ในระบบ)
Habitation_NFM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้นอกระบบ)
Education_NFM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้นอกระบบ)
Consumer_NFM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้นอกระบบ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	คำอธิบาย
Business_NFM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้นอกระบบ)
Farm_NFM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้นอกระบบ)
Other_NFM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหายฯลฯ (หนี้นอกระบบ)
PayOff_Debt	ใน 12 เดือนที่แล้ว มีการชำระหนี้สินทุกประเภทหรือไม่
PD_Land	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินประเภทของการซื้อ/เช่าบ้าน และ/หรือที่ดิน
PD_Education	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินในการศึกษา
PD_Consumer	ได้ชำระหนี้สินประเภทของการใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน

ตารางที่ 3.3 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลด้านสังคม

ชื่อตัวแปร	คำอธิบาย
Members	จำนวนสมาชิกครัวเรือน
MaritalStatus_HF	สถานะสมรสของหัวหน้าครัวเรือน
SEX_HF	เพศของหัวหน้าครัวเรือน
Age_HF	อายุของหัวหน้าครัวเรือน
PWD	จำนวนผู้พิการในแต่ละครัวเรือน
Education	จำนวนผู้ที่กำลังเรียนหนังสืออยู่ในแต่ละครัวเรือน
Welfare	จำนวนสวัสดิการด้านการรักษาพยาบาลในแต่ละครัวเรือนที่ได้รับ
AidGov	จำนวนโครงการช่วยเหลือของรัฐที่เคยได้รับในแต่ละครัวเรือน
LoansGov	จำนวนการกู้ยืมเงินจากกองทุนต่างๆของรัฐในแต่ละครัวเรือน
Earners	จำนวนผู้ทำงานหารรายได้ในครัวเรือน
Child_Elder	จำนวนผู้พิการและเด็กในครัวเรือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 การทำความสะอาดข้อมูล (Data Cleaning)

เป็นกระบวนการตรวจสอบและแก้ไขรายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล ซึ่งเป็นหลักการสำคัญของฐานข้อมูล จึงต้องมีการแทนที่ การปรับปรุง หรือการลบข้อมูลที่ไม่ถูกต้องออกไป เพื่อให้ได้ข้อมูลที่มีคุณภาพ

ตารางที่ 3.4 Python Code สำหรับการหาข้อมูลสูญหาย

```
Microdata_df.isnull().sum()
```

จากตาราง 3.4 เป็นการแสดง Python Code สำหรับการหาข้อมูลสูญหาย หนึ่งในขั้นตอนการทำทำความสะอาดข้อมูล โดยขั้นตอนนี้เป็นการตรวจสอบหาข้อมูลสูญหาย เนื่องจากข้อมูลที่ได้รับการตรวจสอบแล้วไม่พบข้อมูลสูญหาย ดังนั้นสามารถนำข้อมูลไปใช้ในขั้นตอนถัดไป

3.1.3 การแปลงข้อมูล (Data Transformation)

เป็นการเปลี่ยนลักษณะข้อมูลเพื่อให้เหมาะสมกับอัลกอริทึม โดยอัลกอริทึมจะทำงานได้ดีเมื่อข้อมูลอยู่ในรูปแบบคะแนนมาตรฐาน เพื่อให้ข้อมูลพร้อมใช้งานดังนั้นเราต้องทำการแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกัน (Standardization) ด้วยวิธี Z-Score หรือ Standard Scaler โดยสามารถคำนวณได้ตามสมการ โดยมีคำสั่งดังตารางที่ 3.5

$$x_{stand} = \frac{x - \bar{x}}{SD(x)}$$

ตารางที่ 3.5 Python Code สำหรับการแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกัน

```
Household_scaler = StandardScaler()
Household_sc = Household_scaler.fit_transform(Household_df)
Cols = Household_df.columns
Household_sc = pd.DataFrame(Household_sc, columns=[Cols])
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.6 ข้อมูลตัวอย่างของข้อมูลที่อยู่ในรูปมาตรฐานเดียวกัน

NEW_HH_NO	Members	Age_HF	PWD	Education	Welfare
1	-0.22926	0.186427	-0.29423	-0.50098	-0.25434
9	-0.22926	0.003573	-0.29423	-0.50098	-0.25434
16	-0.22926	1.039748	-0.29423	-0.50098	-0.25434
18	0.504687	0.186427	-0.29423	-0.50098	0.397204
19	0.504687	0.30833	-0.29423	0.963472	0.397204

จัดกลุ่มข้อมูลด้วยวิธีเคมีนและดีปีสแกน โดยแยกพิจารณาเป็น 2 กรณีดังนี้

3.1 กรณีที่ครัวเรือนมีหนี้สิน

3.2 กรณีที่ครัวเรือนไม่มีหนี้สิน

3.1.5 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลสำหรับครัวเรือนมีภาระหนี้สิน

ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนครัวเรือนมีภาระหนี้สิน ประกอบด้วยตัวแปรด้านเศรษฐกิจดังตารางที่ 3.7 และตัวแปรด้านสังคมตารางที่ 3.8

ตารางที่ 3.7 ตัวแปรด้านเศรษฐกิจที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปีสแกนของครัวเรือนที่มีภาระหนี้สิน

ชื่อตัวแปร	คำอธิบาย
Habitation_FM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้ในระบบ)
Education_FM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้ในระบบ)
Consumer_FM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้ในระบบ)
Business_FM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้ในระบบ)
Farm_FM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้ในระบบ)
Other_FM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้ำประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย ฯลฯ (หนี้ในระบบ)
Habitation_NFM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้นอกระบบ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้วยการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	คำอธิบาย
Education_NFM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้นอกระบบ)
Consumer_NFM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้นอกระบบ)
Business_NFM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้นอกระบบ)
Farm_NFM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้นอกระบบ)
Other_NFM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย ฯลฯ (หนี้นอกระบบ)
PayOff_Debt	ใน 12 เดือนที่แล้ว มีการชำระหนี้สินทุกประเภทหรือไม่
PD_Land	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินประเภทของการซื้อ/เช่าบ้าน และ/หรือที่ดิน
PD_Education	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินในการศึกษา
PD_Consumer	ได้ชำระหนี้สินประเภทของการใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน
PD_Business	ได้ชำระหนี้สินประเภทของการใช้ในการทำธุรกิจ (ที่ไม่ใช่การเกษตร)
PD_Farm	ได้ชำระหนี้สินประเภทของการใช้ในการทำการเกษตร
PD_Other	ได้ชำระหนี้สินประเภทอื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย
BuyHouse_Saving	ใน 12 เดือนที่แล้ว ครัวเรือนนี้มีการซื้อที่อยู่อาศัย (ด้วยเงินออมของครัวเรือน) หรือไม่
Amount_BuyHouse	จำนวนเงินของครัวเรือนนี้มีการซื้อที่อยู่อาศัย (ด้วยเงินออมของครัวเรือน)
Total_Income	รายได้ทั้งหมดเฉลี่ยต่อเดือน
Total_Cost	ค่าใช้จ่ายทั้งหมดเฉลี่ยต่อเดือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.8 ตัวแปรด้านสังคมที่ใช้ในการแบ่งกลุ่มแบบเคมินและดีปีสแกนของครัวเรือนที่มีภาระหนี้สิน

ชื่อตัวแปร	คำอธิบาย
Members	จำนวนสมาชิกในครัวเรือน
Age_HF	อายุของหัวหน้าครัวเรือน
PWD	จำนวนผู้พิการในแต่ละครัวเรือน
Education	จำนวนผู้ที่กำลังเรียนหนังสืออยู่ในแต่ละครัวเรือน
Welfare	จำนวนสวัสดิการด้านการรักษาพยาบาล
AidGov	จำนวนโครงการช่วยเหลือของรัฐที่เคยได้รับในแต่ละครัวเรือน
LoansGov	จำนวนการกู้ยืมเงินจากกองทุนต่างๆของรัฐในแต่ละครัวเรือน
Earners	จำนวนผู้ทำงานหารายได้ในครัวเรือน
Child_Elder	จำนวนผู้พิการและเด็กในครัวเรือน

3.1.6 ตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลสำหรับครัวเรือนไม่มีภาระหนี้สิน

ในการแบ่งกลุ่มแบบเคมินและดีปีสแกนของครัวเรือนครัวเรือนไม่มีภาระหนี้สิน ประกอบด้วยตัวแปรด้านเศรษฐกิจดังตารางที่ 3.9 และตัวแปรด้านสังคมตารางที่ 3.10

ตารางที่ 3.9 ตัวแปรด้านเศรษฐกิจที่ใช้ในการแบ่งกลุ่มแบบเคมินและดีปีสแกนของครัวเรือนที่ไม่มีภาระหนี้สิน

ชื่อตัวแปร	คำอธิบาย
Habitation_FM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้ในระบบ)
Education_FM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้ในระบบ)
Consumer_FM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้ในระบบ)
Business_FM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้ในระบบ)
Farm_FM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้ในระบบ)
Other_FM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย ฯลฯ (หนี้ในระบบ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	คำอธิบาย
Habitation_NFM	จำนวนหนี้สินเพื่อใช้ซื้อ/เช่าซื้อบ้าน และ/หรือที่ดิน (หนี้นอกระบบ)
Education_NFM	จำนวนหนี้สินเพื่อใช้ในการศึกษา (หนี้นอกระบบ)
Consumer_NFM	จำนวนหนี้สินเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน (หนี้นอกระบบ)
Business_NFM	จำนวนหนี้สินเพื่อใช้ในการทำธุรกิจ(ที่ไม่ใช่การเกษตร) (หนี้นอกระบบ)
Farm_NFM	จำนวนหนี้สินเพื่อใช้ในการทำการเกษตร (หนี้นอกระบบ)
Other_NFM	จำนวนหนี้สินเพื่ออื่นๆ เช่น หนี้จากการค้ำประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย ฯลฯ (หนี้นอกระบบ)
PayOff_Debt	ใน 12 เดือนที่แล้ว มีการชำระหนี้สินทุกประเภทหรือไม่
PD_Land	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินประเภทของการซื้อ/เช่าบ้าน และ/หรือที่ดิน
PD_Education	ใน 12 เดือนที่แล้ว ได้ชำระหนี้สินในการศึกษา
PD_Consumer	ได้ชำระหนี้สินประเภทของการใช้จ่ายอุปโภคบริโภคอื่นๆ ในครัวเรือน
PD_Business	ได้ชำระหนี้สินประเภทของการใช้ในการทำธุรกิจ (ที่ไม่ใช่การเกษตร)
PD_Farm	ได้ชำระหนี้สินประเภทของการใช้ในการทำการเกษตร
PD_Other	ได้ชำระหนี้สินประเภทอื่นๆ เช่น หนี้จากการค้ำประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย
BuyHouse_Saving	ใน 12 เดือนที่แล้ว ครัวเรือนนี้มีการซื้อที่อยู่อาศัย (ด้วยเงินออมของครัวเรือน) หรือไม่
Amount_BuyHouse	จำนวนเงินของครัวเรือนนี้มีการซื้อที่อยู่อาศัย (ด้วยเงินออมของครัวเรือน)
Total_Income	รายได้ทั้งหมดเฉลี่ยต่อเดือน
Total_Cost	ค่าใช้จ่ายทั้งหมดเฉลี่ยเดือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.10 ตัวแปรด้านสังคมที่ใช้ในการแบ่งกลุ่มแบบเคมีนและดีปัสแกนของครัวเรือนที่ไม่มีภาระหนี้สิน

ชื่อตัวแปร	คำอธิบาย
Members	จำนวนสมาชิกในครัวเรือน
Age_HF	อายุของหัวหน้าครัวเรือน
PWD	จำนวนผู้พิการในแต่ละครัวเรือน
Education	จำนวนผู้ที่กำลังเรียนหนังสืออยู่ในแต่ละครัวเรือน
Welfare	จำนวนสวัสดิการด้านการรักษาพยาบาล
AidGov	จำนวนโครงการช่วยเหลือของรัฐที่เคยได้รับในแต่ละครัวเรือน
LoansGov	จำนวนการกู้ยืมเงินจากกองทุนต่างๆของรัฐในแต่ละครัวเรือน
Earners	จำนวนผู้ทำงานหารายได้ในครัวเรือน
Child_Elder	จำนวนผู้พิการและเด็กในครัวเรือน

3.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้คือโปรแกรมภาษา Python และ Power Bi

การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน โดยพิจารณาจำนวนกลุ่มที่เหมาะสมจากวิธี Silhouette Score และการแบ่งกลุ่มข้อมูลด้วยวิธีดีปัสแกน โดยพิจารณาค่า eps ที่เหมาะสมจากวิธี Nearest Neighbors จากนั้นพิจารณาประสิทธิภาพของการแบ่งกลุ่มด้วยวิธี Silhouette Coefficient

3.3 ขั้นตอนวิเคราะห์

3.3.1 ขั้นตอนที่ 1 วิเคราะห์จำนวนกลุ่มที่เหมาะสม

วิธี Silhouette Score

ในการเลือกจำนวนกลุ่มที่เหมาะสมสำหรับการวิเคราะห์การแบ่งกลุ่ม โดยการคำนวณระยะห่างระหว่างกลุ่มเทียบกับระยะห่างระหว่างจุดในกลุ่มเดียวกัน สามารถกำหนดค่าได้จากการเรียกใช้โมดูล silhouette_score จากไลบรารี sklearn.metric ดังตารางที่ 3.11

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.11 Python Code การวิเคราะห์จำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score

```
def sil_score(Household_sc, from_k=2, to_k=7):
    sils=[]
    for k in range(from_k, to_k+1):
        model = KMeans(n_clusters=k,init='kmeans++',random_state=42)
            .fit(Household_sc)
        silhouette_avg = silhouette_score(Household_sc, model.labels_).round(4)
        sils.append([silhouette_avg,k])
    return sils

model = KMeans(n_clusters=..)
Visualizer = SilhouetteVisualizer(model).fit(Household_sc)
Visualizer.poof
```

3.3.2 ขั้นตอนที่ 2 การหาค่าพารามิเตอร์ Eps ที่เหมาะสม

วิธี Nearest Neighbors

การกำหนดค่า eps พิจารณาจากจุดสองจุดหากระยะห่างระหว่างจุดสองจุดต่ำกว่าเกณฑ์จะจัดว่าเป็นเพื่อนบ้านกัน โดยคำนวณระยะห่างจากแต่ละจุดไปยังเพื่อนบ้านที่ใกล้ที่สุดได้จากโมดูล Nearest Neighbors ใช้คำสั่ง n_neighbors ดังตารางที่ 3.12

ตารางที่ 3.12 Python Code การวิเคราะห์หาค่า Eps ที่เหมาะสมด้วยวิธี Nearest Neighbors ครั้งที่ 1

```
neigh = NearestNeighbors(n_neighbors=20)
nbrs = neigh.fit(Household_sc)
distances, indices = nbrs.kneighbors(Household_sc)
distances = np.sort(distances, axis=0)
distances = distances[:,1]
plt.figure(figsize=(10,8))
plt.plot(distances)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

min_samples = range(10,20)
eps = np.arange(2.5,6,1.5)
output = []
for ms in min_samples:
    for ep in eps:
        labels = DBSCAN(min_samples=ms, eps = ep).fit(Household_sc).labels_
        output.append((ms, ep))

min_samples, eps = sorted(output, key=lambda x:x[-1])[-1]
print(f"min_samples: {min_samples}")
print(f"eps: {eps}")

```

ตารางที่ 3.13 Python Code การวิเคราะห์หาค่า Eps ที่เหมาะสมด้วยวิธี Nearest Neighbors ครั้งที่ 2

```

neigh = NearestNeighbors(n_neighbors=20)
nbrs = neigh.fit(Household_sc)
distances, indices = nbrs.kneighbors(Household_sc)
distances = np.sort(distances, axis=0)
distances = distances[:,1]
plt.figure(figsize=(10,8))
plt.plot(distances)

min_samples = range(10,100)
eps = np.arange(2.5,6,1.5)
output = []
for ms in min_samples:
    for ep in eps:
        labels = DBSCAN(min_samples=ms, eps = ep).fit(Household_sc).labels_
        output.append((ms, ep))

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
min_samples, eps = sorted(output, key=lambda x:x[-1])[-1]
print(f"min_samples: {min_samples}")
print(f"eps: {eps}")
```

3.3.3 ขั้นตอนที่ 3 การแบ่งกลุ่ม (Clustering)

วิธีเคมีน (K-means clustering)

จัดกลุ่มข้อมูลโดยใช้วิธีเคมีน ซึ่งกำหนดค่าในกลุ่มของครัวเรือนไม่มีหนี้สิน $K=7, 2$ และกลุ่มครัวเรือนที่มีหนี้สิน $K=5, 2$ ที่ได้จากการหาจำนวนกลุ่มที่เหมาะสม ด้วยวิธี Elbow Method และ Silhouette Score โดยใช้โมดูล KMeans จากไลบรารี sklearn.cluster ดังตารางที่ 3.14

ตารางที่ 3.14 Python Code การแบ่งกลุ่มโดยวิธีเคมีน

```
kmeans = KMeans(n_clusters = 2 ;init='kmeans++',random_state=42).fit(Household_sc)
identified_cluster1 = kfit.predict(Household_sc)
```

วิธีดีบีเอสแกน (DBSCAN clustering)

จัดกลุ่มข้อมูลโดยใช้วิธีดีบีเอสแกน ซึ่งกำหนดค่าในกลุ่มของครัวเรือนไม่มีหนี้สิน $eps = 2.9$, $MinPts = 19$ และกลุ่มครัวเรือนที่มีหนี้สิน $eps = 5.5$, $MinPts = 19$ ใช้โมดูล DBSCAN จากไลบรารี sklearn.cluster ดังตารางที่ 3.15

ตารางที่ 3.15 Python Code การแบ่งกลุ่มโดยวิธีดีบีเอสแกน ครั้งที่ 1

```
labels = DBSCAN(min_samples = 19, eps = 5.5).fit(Household_sc).labels_
clusters = len(Counter(labels))
print(f"Number of clusters: {clusters}")
print(f"Number of outliers: {Counter(labels)[-1]}")
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.16 Python Code การแบ่งกลุ่มโดยวิธีดีปีสแกน ครั้งที่ 2

```
labels = DBSCAN(min_samples = 99, eps = 5.5).fit(Household_sc).labels_
clusters = len(Counter(labels))
print(f"Number of clusters: {clusters}")
print(f"Number of outliers: {Counter(labels)[-1]}")
```

3.3.4 ขั้นตอนที่ 4 การวัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี Silhouette Coefficient

วิธี Silhouette Coefficient เป็นค่าณระยะห่างระหว่างกลุ่มเทียบกับระยะห่างระหว่างจุดในกลุ่มเดียวกัน โดยนำข้อมูลที่ถูกรวบรวมด้วยวิธีเคมีนและดีปีสแกนมาวิเคราะห์ เรียกใช้โมดูล metrics จากไลบรารี sklearn ดังตารางที่ 3.17

ตารางที่ 3.17 Python Code การวัดประสิทธิภาพการแบ่งกลุ่มด้วยวิธี Silhouette Coefficient

```
print(f"Silhouette Coefficient: {metrics.silhouette_score(Household_sc,
                                                         identified_cluster):.3f}")
print(f"Silhouette Coefficient: {metrics.silhouette_score(Household_sc, labels):.3f}")
```

บทที่ 4

ผลการวิจัย

จากการวิเคราะห์ข้อมูลในการศึกษาภาระหนี้สินของครัวเรือนในประเทศไทย โดยวิเคราะห์ข้อมูลทุติยภูมิที่ได้ทำการสำรวจข้อมูลภาวะเศรษฐกิจและครัวเรือน ปี พ.ศ. 2564 ที่วราชนาจนจักร โดยสำนักงานสถิติแห่งชาติมีรายละเอียดดังต่อไปนี้

- 4.1 การวิเคราะห์ลักษณะทั่วไปและสัดส่วนของครัวเรือนไทยที่มีหนี้และไม่มีหนี้สิน
- 4.2 การวิเคราะห์แบ่งกลุ่มครัวเรือนมีภาระหนี้สิน
 - 4.2.2 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมิน
 - 4.2.2 การแบ่งกลุ่มข้อมูลด้วยวิธีดีปีสแกน
- 4.3 การวิเคราะห์แบ่งกลุ่มครัวเรือนไม่มีภาระหนี้สิน
 - 4.3.2 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมิน
 - 4.3.2 การแบ่งกลุ่มข้อมูลด้วยวิธีดีปีสแกน
- 4.4 การวิเคราะห์เปรียบเทียบการแบ่งกลุ่มด้วย Silhouette Coefficient
- 4.5 การทดสอบความแตกต่างค่าเฉลี่ย
- 4.6 แดชบอร์ด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1 การวิเคราะห์ลักษณะทั่วไปและสัดส่วนของครัวเรือนไทยที่มีภาระหนี้และไม่มีภาระหนี้สิน

1) ลักษณะทางภูมิศาสตร์และที่อยู่อาศัยของครัวเรือนในประเทศไทย

เมื่อพิจารณาตามลักษณะที่ตั้งภูมิศาสตร์ของครัวเรือนในประเทศไทย ปี พ.ศ. 2565 พบว่าครัวเรือนส่วนใหญ่กระจายอยู่ในภาคกลาง รองลงมาคือภาคตะวันออกเฉียงเหนือ ภาคเหนือ ภาคใต้ และกรุงเทพมหานครและปริมณฑล โดยมีรายละเอียดดังนี้ ภาคตะวันออกเฉียงเหนือร้อยละ 26.5 จำแนกเป็นครัวเรือนที่มีหนี้ถึงร้อยละ 16.5 และ รองลงมาคือภาคกลางร้อยละ 13.8 ภาคเหนือร้อยละ 12.2 ภาคใต้ร้อยละ 7.3 และ กรุงเทพมหานครและปริมณฑลร้อยละ 1.9 ตามลำดับ และส่วนใหญ่เป็นครัวเรือนที่อาศัยอยู่ในเขตเทศบาลร้อยละ 56.6 จำแนกเป็นครัวเรือนที่มีหนี้สิน 27.5 ดังตารางที่ 4.1

ตารางที่ 4.1 ร้อยละของครัวเรือนในประเทศไทย จำแนกตามภาค เขตปกครอง

ภาคและเขตปกครอง	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
1. ภาค			
กรุงเทพมหานครและปริมณฑล *	5.6	1.9	3.7
ภาคกลาง	29.2	13.8	15.4
ภาคเหนือ	23.1	12.2	10.9
ภาคตะวันออกเฉียงเหนือ	26.5	16.5	10.0
ภาคใต้	15.6	7.3	8.3
2. เขตการปกครอง			
ในเขตเทศบาล	56.6	27.5	29.1
นอกเขตเทศบาล	43.4	24.3	19.1

*หมายถึง สมุทรปราการ นนทบุรี ปทุมธานี นครปฐม สมุทรสาคร

เมื่อพิจารณาตามลักษณะที่อยู่อาศัยของครัวเรือนในประเทศไทย พบว่าครัวเรือนส่วนใหญ่มีลักษณะที่อยู่อาศัยเป็นประเภทบ้านเดี่ยวถึงร้อยละ 80.4 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 43.4 หากพิจารณาตามสถานภาพการครอบครองที่อาศัย พบว่าส่วนใหญ่เป็นเจ้าของบ้านและที่ดินร้อยละ 76.5 จำแนกเป็นครัวเรือนที่มีหนี้สินร้อยละ 41.2 ดังตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ร้อยละของครัวเรือนในประเทศไทย จำแนกลักษณะที่อยู่อาศัย

ลักษณะที่อยู่อาศัย	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
1. ประเภทที่อยู่อาศัย			
บ้านเดี่ยว	80.4	43.4	37.0
ห้องแถว/ตึกแถว/อาคารพาณิชย์	12.2	4.9	7.3
ทาวน์เฮาส์/บ้านแฝด	4.2	2.2	2.0
ห้องชุด	2.8	1.1	1.7
ห้องภายในบ้าน	0.3	0.1	0.2
ที่อยู่ชั่วคราว/อื่นๆ	0.1	0.0	0.1
2. สถานภาพครอบครองที่อยู่อาศัย			
เป็นเจ้าของบ้านและที่ดิน	76.5	41.2	35.3
เป็นเจ้าของบ้านแต่เช่าที่ดิน	1.6	0.7	0.9
เป็นเจ้าของบ้านบนที่ดินสาธารณะ	1.5	0.7	0.8
เช่าซื้อ	0.8	0.8	0.0
เช่า	12.7	5.3	7.4
อยู่โดยไม่เสียค่าเช่า	6.9	3.0	3.9

2) ลักษณะประชากรของหัวหน้าครัวเรือนในประเทศไทย

เมื่อพิจารณาตามลักษณะทางประชากรของหัวหน้าครัวเรือน พบว่าส่วนใหญ่เป็นเพศชายร้อยละ 57.3 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 31 และเพศหญิงร้อยละ 20.7 ซึ่งเมื่อพิจารณาตามอายุ พบว่าส่วนใหญ่มีอายุอยู่ในช่วง 60 ปีขึ้นไปร้อยละ 42.2 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 18.2 รองลงมาคือมีอายุอยู่ในช่วง 50 ถึง 59 ปีร้อยละ 25.9 อายุ 40 ถึง 49 ปีร้อยละ 17.8 อายุ 30 ถึง 39 ปีร้อยละ 9.6 ตามลำดับ สถานภาพสมรสของหัวหน้าครัวเรือน พบว่าส่วนใหญ่หัวหน้าครัวเรือน มีสถานภาพสมสร้อยละ 61.2 จำแนกเป็นครัวเรือนที่มีหนี้ถึงร้อยละ 36.8 รองลงมาคือ ม่ายร้อยละ 19.6 โสดร้อยละ 11.1 แยกกันอยู่ร้อยละ 4.1 และหย่าร้างร้อยละ 4.0 ระดับการศึกษาของหัวหน้า ครัวเรือน พบว่าส่วนใหญ่หัวหน้าครัวเรือนเกินครึ่งมีระดับการศึกษาประถมศึกษาและ ต่ำกว่าร้อยละ 55.6 จำแนกเป็นครัวเรือนที่มี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หนี้ร้อยละ 28.3 รองลงมาคือ มัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลายร้อยละ 23.5 ปริญญาตรี ร้อยละ 10.6 ดังตารางที่ 4.3

ตารางที่ 4.3 ร้อยละของครัวเรือนในประเทศไทย จำแนกลักษณะทางประชากรของหัวหน้าครัวเรือน

ลักษณะหัวหน้าครัวเรือน	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
1. เพศ			
ชาย	57.3	31.0	26.3
หญิง	42.7	20.7	22.0
2. อายุ			
ต่ำกว่า 20 ปี	0.5	0.0	0.4
20 – 29 ปี	3.1	1.5	2.6
30 – 39 ปี	9.6	5.4	4.2
40 – 49 ปี	17.8	10.7	7.1
50 – 59 ปี	25.9	15.9	10.0
60 ปี ขึ้นไป	42.2	18.2	24.0
3. สถานภาพสมรส			
โสด	11.1	3.5	7.6
สมรส	61.2	36.8	24.4
ม่าย	19.6	7.7	11.9
หย่า	4.0	2.0	2.0
แยกกันอยู่	4.1	1.7	2.4
4. ระดับการศึกษาสูงสุด			
ไม่เคยเรียน	4.7	1.4	3.3
ก่อนประถมศึกษาและประถมศึกษา	55.6	28.3	27.3
มัธยมศึกษาตอนต้นและตอนปลาย	23.5	13.0	10.4
อาชีวศึกษาและอนุปริญญา	3.8	2.2	1.6
ปริญญาตรี	10.6	5.8	4.8
สูงกว่าปริญญาตรี	1.6	1.0	0.6
การศึกษาอื่นๆ	0.2	0.0	0.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) ลักษณะทางประชากรของครัวเรือนในประเทศไทย

เมื่อพิจารณาตามลักษณะทางประชากรของครัวเรือน พบว่าครัวเรือนครึ่งหนึ่งมีจำนวน สมาชิก 1 ถึง 2 คน ร้อยละ 51 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 19.4 รองลงมาคือ จำนวนสมาชิก 3 ถึง 4 คน ร้อยละ 35.3 จำนวนสมาชิก 5 ถึง 7 คน ร้อยละ 12.9 และจำนวนสมาชิก 8 คนขึ้นไปร้อยละ 0.8 ตามลำดับ พิจารณาตามจำนวนผู้ทำงานหารรายได้ในครัวเรือน พบว่าครัวเรือนส่วนใหญ่มีจำนวนผู้ทำงานหารรายได้ 2 ถึง 3 คน ร้อยละ 47.8 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 31.9

หากพิจารณาตามสมาชิกในครัวเรือนที่อยู่ในภาวะพึ่งพิงที่ต้องการได้รับความช่วยเหลือ พบว่าครัวเรือนมีสมาชิกอยู่ในวัยเด็กและวัยสูงอายุอย่างน้อย 1 คน ร้อยละ 50.6 จำแนกเป็น ครัวเรือนที่มีหนี้ถึงร้อยละ 26.9 และพบว่าครัวเรือนมีสมาชิกเป็นผู้พิการอย่างน้อย 1 คนมีเพียงร้อยละ 9.1 จำแนกเป็นครัวเรือนที่มีหนี้เพียงร้อยละ 4.6 ดังตารางที่ 4.4

ตารางที่ 4.4 ร้อยละของครัวเรือนในประเทศไทย จำแนกลักษณะทางประชากรของครัวเรือน

ลักษณะครัวเรือน	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
1. จำนวนสมาชิกในครัวเรือน			
1 – 2 คน	51	19.4	31.6
3 – 4 คน	35.3	22.4	12.9
5 – 7 คน	12.9	9.3	3.6
7 คนขึ้นไป	0.8	0.6	0.2
2. จำนวนผู้มีรายได้ในครัวเรือน			
ไม่มีผู้มีรายได้	15.1	2.3	12.8
1 คน	32.5	13.9	18.6
2 – 3 คน	47.8	31.9	15.9
4 คนขึ้นไป	4.6	3.6	1.0
3. ครัวเรือนมีสมาชิกอยู่ในวัยเด็กหรือสูงอายุ			
มี	50.6	26.9	23.7
ไม่มี	49.4	24.8	24.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะครัวเรือน	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
4. ครัวเรือนมีสมาชิกเป็นผู้พิการ			
มี	9.1	4.6	4.5
ไม่มี	90.9	47.1	43.8

*หมายเหตุ : วัยเด็ก คือ เด็กอายุต่ำกว่า 15 ปี และวัยสูงอายุคือ อายุ 60 ปีขึ้นไป

4) ลักษณะทางเศรษฐกิจของครัวเรือนในประเทศไทย

เมื่อพิจารณาตามลักษณะทางเศรษฐกิจทั่วไปของครัวเรือน พบว่าครัวเรือนมีรายได้ทั้งสิ้น เฉลี่ยเดือนละ 28,149 บาท ซึ่งส่วนใหญ่มีรายได้ทั้งสิ้นอยู่ในช่วง 15,001 ถึง 30,000 บาทต่อเดือน ร้อยละ 37.6 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 20.7 รองลงมาคือ รายได้ทั้งสิ้นอยู่ในช่วง 10,001 ถึง 15,000 บาทต่อเดือนร้อยละ 18.8 ส่วนครัวเรือนที่มีรายได้ทั้งสิ้นช่วงต่ำกว่าหรือเท่ากับ 5,000 บาทต่อเดือนมีเพียงร้อยละ 1.5 และครัวเรือนที่มีรายได้ทั้งสิ้นสูงกว่า 30,000 บาทต่อเดือนขึ้นไป มีสูงถึงร้อยละ 29.5

ในส่วนของค่าใช้จ่ายทั้งสิ้นเฉลี่ยต่อเดือนของครัวเรือน พบว่าครัวเรือนมีค่าใช้จ่ายทั้งสิ้น เฉลี่ยต่อเดือนละ 15,956 บาท ซึ่งส่วนใหญ่มีค่าใช้จ่ายทั้งสิ้นอยู่ในช่วง 15,001 ถึง 30,000 บาทต่อเดือนร้อยละ 26.6 จำแนกเป็นครัวเรือนที่มีหนี้ร้อยละ 16.6 รองลงมาคือ ค่าใช้จ่ายทั้งสิ้นอยู่ในช่วง 5,001 ถึง 10,000 บาทต่อเดือนร้อยละ 26.4 และครัวเรือนที่มีค่าใช้จ่ายทั้งสิ้นอยู่ในช่วง 10,001 ถึง 15,000 บาทต่อเดือนร้อยละ 20.1 ส่วนครัวเรือนที่มีค่าใช้จ่ายทั้งสิ้นต่ำกว่าหรือเท่ากับ 5,000 บาทต่อเดือนมีเพียงร้อยละ 15 และหากพิจารณาครัวเรือนที่มีค่าใช้จ่ายทั้งสิ้น สูงกว่า 30,000 บาทต่อเดือนขึ้นไปมีร้อยละ 11.9

ตารางที่ 4.5 ร้อยละของครัวเรือนในประเทศไทย จำแนกลักษณะทางเศรษฐกิจทั่วไปของครัวเรือน

ลักษณะทางเศรษฐกิจด้านการเงิน ของครัวเรือน	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
1. รายได้ทั้งสิ้นเฉลี่ยต่อเดือน			
ต่ำกว่า 1,500 บาท	0.2	0.1	0.1
1,500 – 3,000 บาท	0.1	0.0	0.1
3,001 – 5,000 บาท	1.2	0.2	1.0
5,001 - 10,000 บาท	12.6	3.0	9.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับกรใช้งานเพื่อการศึกษาเท่านั้น ไมอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะทางเศรษฐกิจด้านการเงิน ของครัวเรือน	ร้อยละของครัวเรือน		
	ทั้งสิ้น	มีหนี้	ไม่มีหนี้
ทั่วราชอาณาจักร	100.0	51.7	48.3
10,001 – 15,000 บาท	18.8	7.4	11.4
15,001 – 30,000	37.6	20.7	16.9
30,001 - 50,000	17.9	12.1	5.8
50,001 – 100,000	9.4	6.7	2.7
มากกว่า 100,000 บาท	2.2	1.5	0.7
2. ค่าใช้จ่ายทั้งสิ้นเฉลี่ยต่อเดือน			
ต่ำกว่า 1,500 บาท	1.7	0.1	1.6
1,500 – 3,000 บาท	4.6	0.7	3.9
3,001 – 5,000 บาท	8.7	2.8	5.9
5,001 - 10,000 บาท	26.4	12.1	14.3
10,001 – 15,000 บาท	20.1	10.4	9.7
15,001 – 30,000	26.6	16.6	10.0
30,001 - 50,000	8.9	6.7	2.2
50,001 – 100,000	2.7	2.1	0.6
มากกว่า 100,000 บาท	0.3	0.2	0.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

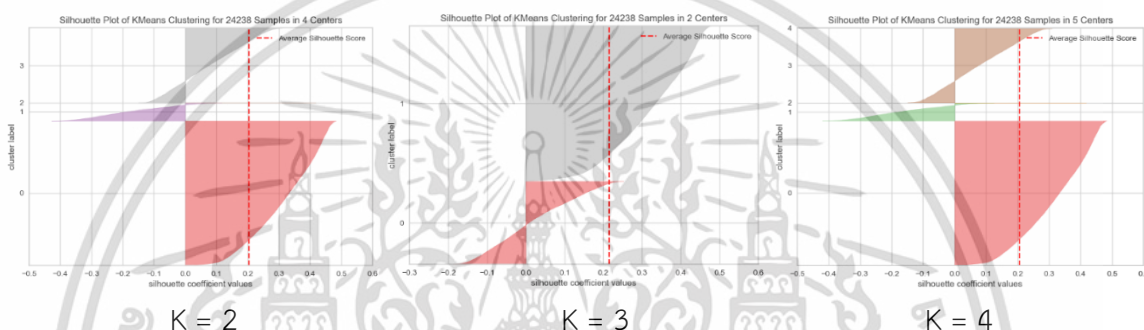
4.2 การวิเคราะห์แบ่งกลุ่มคร้วเรือนมีภาระหนี้สิน

การวิเคราะห์ข้อมูลสำหรับคร้วเรือนมีภาระหนี้สินจำนวน 24,238 คร้วเรือน

4.2.1 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน

Silhouette Score

จากรูปที่ 4.1 เมื่อพิจารณาจากกราฟแสดงผล Silhouette Plot พบว่าไม่สามารถเลือกจำนวนกลุ่มที่เหมาะสมได้เนื่องจากทุกกราฟมีกลุ่มที่ต่ำกว่าค่าเฉลี่ยของ Silhouette score (เส้นประแนวตั้ง) และมีความกว้างของกลุ่มที่แตกต่างกันมาก จึงต้องพิจารณาจากค่า Silhouette score ดังตารางที่ 4.6



รูปที่ 4.1 กราฟการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) เท่ากับ 2 3 และ 4 ด้วยวิธี Silhouette Score

ตารางที่ 4.6 ผลการวิเคราะห์หาจำนวนกลุ่มที่เหมาะสม (K) คร้วเรือนมีภาระหนี้สินด้วยวิธี Silhouette Score

กลุ่ม (K)	Silhouette Score
2	0.2154
3	0.2032
4	0.2052
5	0.2058
6	0.102
7	0.2086

เมื่อพิจารณาผลการวิเคราะห์หาค่า K ที่เหมาะสมจากตารางที่ 4.6 พบว่าค่า K ที่เหมาะสมที่สุดมีค่าเท่ากับ 2 ดังนั้น จำนวนกลุ่มที่เหมาะสมในการแบ่งกลุ่มคร้วเรือนมีภาระหนี้สิน คือ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการวิเคราะห์เคมีน

จากข้อมูลกลุ่มครัวเรือนที่มีภาระหนี้สิน ได้ทำการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score โดยจำนวนกลุ่มที่เหมาะสมเท่ากับ 2 และได้ทำการแบ่งกลุ่มด้วยวิธีเคมีน ผลการวิเคราะห์เป็นดังตาราง 4.7

ตารางที่ 4.7 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมีนครัวเรือนมีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนสมาชิกของครัวเรือน	4.93	2.33
อายุหัวหน้าครัวเรือน	56.17	53.68
จำนวนผู้พิการในครัวเรือน	0.16	0.08
จำนวนผู้ที่กำลังศึกษาในครัวเรือน	1.44	0.26
จำนวนสวัสดิการด้านการรักษาพยาบาล	5.31	2.43
จำนวนโครงการช่วยเหลือจากรัฐบาล	4.55	2.00
จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล	0.50	0.41
จำนวนผู้หารายได้ของครัวเรือน	2.51	1.65
จำนวนของผู้สูงอายุและเด็กในครัวเรือน	1.74	0.36
(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	137,969.62	119,484.45
(หนี้ในระบบ)จำนวนหนี้สินของการศึกษา	7,565.12	6,634.98
(หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	195,163.94	135,843.66
(หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	73,294.25	30,100.99
(หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร	91,779.42	47,170.94
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	3,794.54	2,490.97
(หนี้นอกระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	887.50	794.62
(หนี้นอกระบบ)จำนวนหนี้สินของการศึกษา	85.86	117.58
(หนี้นอกระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	3,334.52	2,409.13
(หนี้นอกระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	2,525.63	1,376.20
(หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร	1,186.20	467.88
(หนี้นอกระบบ)จำนวนหนี้สินอื่นๆ	111.72	353.58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย	1,322.85	1,267.89
จำนวนเงินชำระหนี้การศึกษา	79.89	90.68
จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ	4,758.47	3,398.00
จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร	1,117.11	518.25
จำนวนเงินชำระหนี้การทำเกษตร	1,283.70	733.16
จำนวนเงินชำระหนี้อื่นๆ	25.09	23.06
จำนวนเงินที่ซื้อที่อยู่อาศัย	67.76	33.51
จำนวนรายได้เฉลี่ยต่อเดือน	42,051.72	28,420.23
จำนวนรายจ่ายเฉลี่ยต่อเดือน	24,308.40	16,679.11

กลุ่มที่ 1 (Cluster 1)

ประชากรในกลุ่มที่ 3 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 5 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 6 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 3 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 3 คน มีผู้พิการและเด็กในครัวเรือน 2 คน และมีผู้ที่กำลังศึกษา 1 คน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 42,051 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 24,308 บาท ครัวเรือนมีวัตถุประสงค์ในการกู้ยืมที่มีจำนวนสูงสุด คือการกู้ยืมเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆในครัวเรือน โดยมีค่าเฉลี่ยอยู่ที่ 195,163 บาท และได้ทำการชำระหนี้ในส่วนของการกู้ยืมเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆในครัวเรือนอยู่ที่ 4,758 บาทต่อเดือน

กลุ่มที่ 2 (Cluster 2)

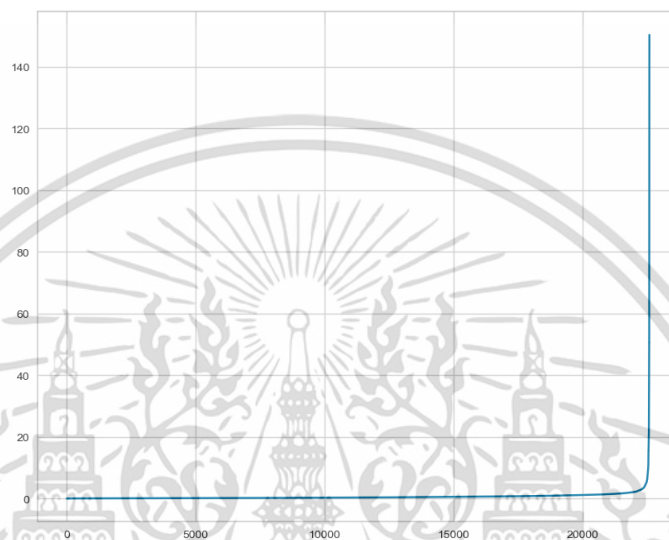
ประชากรในกลุ่มที่ 2 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 2 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 2 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 2 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 2 คน ไม่มีผู้พิการ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 28,420 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 16,679 บาท ครัวเรือนมีวัตถุประสงค์ในการกู้ยืมที่มีจำนวนสูงสุด คือการกู้ยืมเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆในครัวเรือน โดยมีค่าเฉลี่ยอยู่ที่ 135,843 บาท และได้มีการชำระหนี้ในส่วนของการกู้ยืมเพื่อใช้จ่ายอุปโภคบริโภคอื่นๆในครัวเรือนอยู่ที่ 3,398 บาทต่อเดือน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 การแบ่งกลุ่มข้อมูลด้วยวิธีทีบีสแกน

Nearest Neighbor

ผลของการค้นหาช่วงของ eps ที่เหมาะสมโดยวิธี Nearest Neighbor แสดงเป็นรูปแบบกราฟดังต่อไปนี้



รูปที่ 4.2 กราฟหาค่า eps ที่เหมาะสมด้วยวิธี Nearest Neighbor

จากรูปที่ 4.2 จะพิจารณาจุดหักศอกของกราฟ k-distance พบว่ามีค่าประมาณที่ 2 ถึง 5 ดังนั้นผู้วิจัยจึงประมาณค่าพารามิเตอร์ลงในขั้นตอนการวิเคราะห์หาค่า eps ที่เหมาะสม และผลของค่า eps ที่เหมาะสมที่สุดมีค่าเท่ากับ 5.5

ผลการวิเคราะห์ทีบีสแกน

จากข้อมูลกลุ่มครัวเรือนที่มีภาระหนี้สินได้ทำการหาค่า min_samples และ eps ที่เหมาะสมด้วยวิธี Nearest Neighbor จากนั้นดำเนินการแบ่งกลุ่มด้วยทีบีสแกน ครั้งที่ 1 กำหนดให้ min_samples เท่ากับ 19 และ eps เท่ากับ 5.5 ครั้งที่ 2 กำหนดให้ min_samples เท่ากับ 99 และ eps เท่ากับ 5.5 จากผลการวิเคราะห์ทั้ง 2 ครั้งพบว่าไม่สามารถแบ่งกลุ่มข้อมูลได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

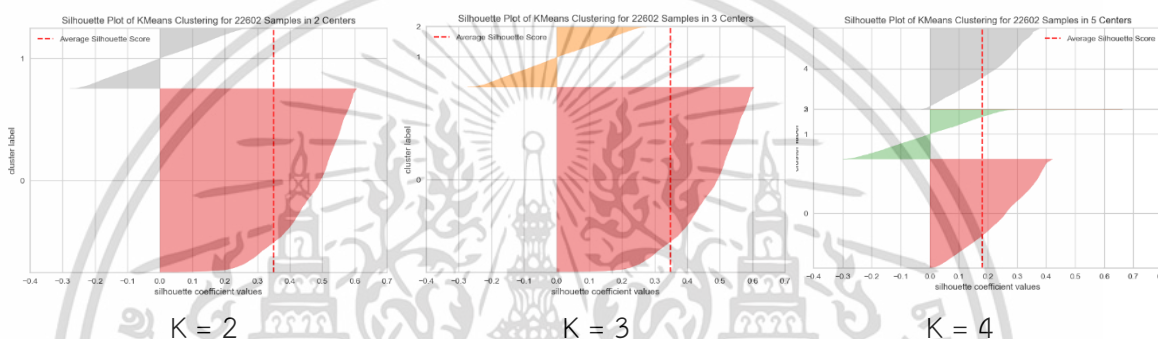
4.3 การวิเคราะห์การแบ่งกลุ่มครวเรือนไม่มีภาระหนี้สิน

การวิเคราะห์ข้อมูลสำหรับครวเรือนไม่มีภาระหนี้สินจำนวน 22,602 ครวเรือน

4.3.1 การแบ่งกลุ่มข้อมูลด้วยวิธีเคมีน

Silhouette Score

จากรูปที่ 4.3 เมื่อพิจารณาจากกราฟแสดงผล Silhouette Plot พบว่าไม่สามารถเลือกจำนวนกลุ่มที่เหมาะสมได้เนื่องจากทุกกราฟมีกลุ่มที่ต่ำกว่าค่าเฉลี่ยของ Silhouette score (เส้นประแนวตั้ง) และมีความกว้างของกลุ่มที่แตกต่างกันมาก จึงต้องพิจารณาจากค่า Silhouette score ดังตารางที่ 4.8



รูปที่ 4.3 กราฟการหาจำนวนกลุ่มที่เหมาะสมที่สุด (K) เท่ากับ 2 3 และ 4 ด้วยวิธี Silhouette Score

ตารางที่ 4.8 ผลการวิเคราะห์หาค่า K ที่เหมาะสมครวเรือนไม่มีภาระหนี้สินด้วยวิธี Silhouette Score

กลุ่ม (K)	Silhouette Score
2	0.3497
3	0.349
4	0.187
5	0.180
6	0.187
7	0.192

เมื่อพิจารณาผลการวิเคราะห์หาค่า K ที่เหมาะสมจากตารางที่ 4.8 พบว่าค่า K ที่เหมาะสมที่สุดมีค่าเท่ากับ 2 ดังนั้น จำนวนกลุ่มที่เหมาะสมในการแบ่งกลุ่มครวเรือนไม่มีภาระหนี้สิน คือ 2 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ผลการวิเคราะห์เคมีน

จากข้อมูลกลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน ได้ทำการหาจำนวนกลุ่มที่เหมาะสมด้วยวิธี Silhouette Score โดยจำนวนกลุ่มที่เหมาะสมเท่ากับ 2 และได้ทำการแบ่งกลุ่มด้วยวิธีเคมีน ผลการวิเคราะห์เป็นดังตารางที่ 4.9

ตารางที่ 4.9 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมีนครัวเรือนไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนสมาชิกของครัวเรือน	1.70	4.19
อายุหัวหน้าครัวเรือน	57.93	57.98
จำนวนผู้พิการในครัวเรือน	0.09	0.15
จำนวนผู้ที่กำลังศึกษาในครัวเรือน	0.09	1.13
จำนวนสวัสดิการด้านการรักษาพยาบาล	1.72	4.46
จำนวนโครงการช่วยเหลือจากรัฐบาล	1.62	3.68
จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล	0.00	0.01
จำนวนผู้หารายได้ของครัวเรือน	0.94	1.95
จำนวนของผู้สูงอายุและเด็กในครัวเรือน	0.44	1.49
(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	49.32	477.48
(หนี้ในระบบ)จำนวนหนี้สินของการศึกษา	26.39	291.78
(หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	508.40	2,011.49
(หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	81.87	183.68
(หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร	110.66	467.00
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	49.50	180.18
(หนี้นอกระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	0.00	0.72
(หนี้นอกระบบ)จำนวนหนี้สินของการศึกษา	0.00	15.32
(หนี้นอกระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	40.01	94.86
(หนี้นอกระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	18.53	16.22
(หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร	31.26	13.51
(หนี้นอกระบบ)จำนวนหนี้สินอื่นๆ	15.19	0.00
จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย	3.01	18.83

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ทำซ้ำโดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนเงินชำระหนี้การศึกษา	0.22	0.57
จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ	10.52	63.59
จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร	1.31	1.28
จำนวนเงินชำระหนี้การทำเกษตร	3.48	19.29
จำนวนเงินชำระหนี้อื่นๆ	0.00	0.00
จำนวนเงินที่ซื้อที่อยู่อาศัย	2.80	0.00
จำนวนรายได้เฉลี่ยต่อเดือน	18,330.69	36,348.17
จำนวนรายจ่ายเฉลี่ยต่อเดือน	10,158.71	18,973.61

กลุ่มที่ 1 (Cluster 1)

ประชากรในกลุ่มที่ 1 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 2 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 2 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 2 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 1 คน ไม่มีผู้พิการ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 18,331 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 10,159 บาท

กลุ่มที่ 2 (Cluster 2)

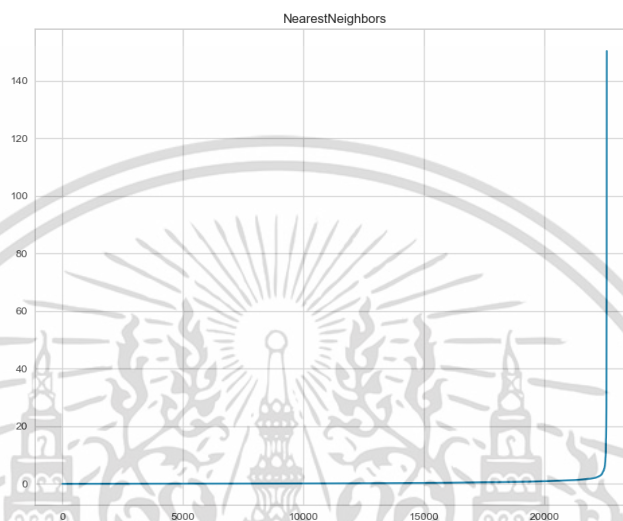
ประชากรในกลุ่มที่ 2 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 4 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 4 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 4 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 2 คน มีผู้พิการและเด็กในครัวเรือน 1 คน และมีผู้ที่กำลังศึกษา 1 คน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 36,348 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 18,974 บาท

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 การแบ่งกลุ่มข้อมูลด้วยวิธีทีบีสแกน

Nearest Neighbor

ผลของการค้นหาช่วงของ eps ที่เหมาะสมโดยวิธี Nearest Neighbor แสดงเป็นรูปแบบกราฟดังต่อไปนี้



รูปที่ 4.4 กราฟหาค่า eps ที่เหมาะสมด้วยวิธี Nearest Neighbor

จากรูปที่ 4.4 จะพิจารณาจุดหักศอกของกราฟ k-distance พบว่ามีค่าประมาณที่ 2 ถึง 5 ดังนั้นผู้วิจัยจึงประมาณค่าพารามิเตอร์ลงในขั้นตอนการวิเคราะห์หาค่า eps ที่เหมาะสม และผลของค่า eps ที่เหมาะสมที่สุดมีค่าเท่ากับ 5.5

ผลการวิเคราะห์ทีบีสแกน

จากข้อมูลกลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน ได้ทำการหาค่า eps ที่เหมาะสมด้วยวิธี Nearest Neighbor ซึ่งค่า eps ที่เหมาะสมอยู่ที่ 5.5 และได้ทำการแบ่งกลุ่มด้วยวิธีทีบีสแกน จากผลการวิเคราะห์สามารถแบ่งกลุ่มข้อมูลได้ 2 กลุ่ม ดังตาราง 4.10

ตารางที่ 4.10 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีตีปีสแกนของครัวเรือนไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนสมาชิกของครัวเรือน	2.30	2.89
อายุหัวหน้าครัวเรือน	57.95	59.00
จำนวนผู้พิการในครัวเรือน	0.11	0.07
จำนวนผู้ที่กำลังศึกษาในครัวเรือน	0.34	0.71
จำนวนสวัสดิการด้านการรักษาพยาบาล	2.38	2.93
จำนวนโครงการช่วยเหลือจากรัฐบาล	2.12	2.93
จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล	0.00	1.00
จำนวนผู้หารายได้ของครัวเรือน	1.18	1.64
จำนวนของผู้สูงอายุและเด็กในครัวเรือน	0.70	0.82
(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	4.25	0.00
(หนี้ในระบบ)จำนวนหนี้สินของการศึกษา	0.89	0.00
(หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	270.08	25,185.71
(หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	8.89	5,357.14
(หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร	37.81	0.00
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	0.00	0.00
(หนี้นอกระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	0.00	0.00
(หนี้นอกระบบ)จำนวนหนี้สินของการศึกษา	0.00	0.00
(หนี้นอกระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	6.30	142.86
(หนี้นอกระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	0.72	0.00
(หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร	0.27	0.00
(หนี้นอกระบบ)จำนวนหนี้สินอื่นๆ	0.40	0.00
จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย	0.07	0.00
จำนวนเงินชำระหนี้การศึกษา	0.00	0.00
จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ	8.59	63.11
จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร	0.00	0.00
จำนวนเงินชำระหนี้ทำการเกษตร	1.86	0.00
จำนวนเงินชำระหนี้อื่นๆ	0.00	0.00

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	Cluster 1	Cluster 2
จำนวนเงินที่ซื้อที่อยู่อาศัย	0.00	0.00
จำนวนรายได้เฉลี่ยต่อเดือน	22,103.73	17,949.25
จำนวนรายจ่ายเฉลี่ยต่อเดือน	12,183.18	10,186.35

กลุ่มที่ 1 (Cluster 1)

ประชากรในกลุ่มที่ 1 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 2 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 2 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 2 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 1 คน มีผู้พิการและเด็กในครัวเรือน 1 คน และไม่มีผู้ที่กำลังศึกษาในครัวเรือน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 22,104 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 12,183 บาท

กลุ่มที่ 2 (Cluster 2)

ประชากรในกลุ่มที่ 2 มีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 3 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50-59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 3 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 3 โครงการ และมีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ 1 กองทุน จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 2 คน มีผู้พิการและเด็กในครัวเรือน 1 คน และมีผู้ที่กำลังศึกษา 1 คน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 17,949 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 10,186 บาท

4.4 การวิเคราะห์เปรียบเทียบการแบ่งกลุ่มด้วย Silhouette Coefficient

ตารางที่ 4.11 การเปรียบเทียบการแบ่งกลุ่มของครัวเรือนที่มีภาระหนี้สิน

วิธีการแบ่งกลุ่ม	Silhouette Coefficient
ครัวเรือนที่มีภาระหนี้สิน ซึ่งแบ่งกลุ่มด้วยวิธีเคมีน โดยหาจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2 ด้วยวิธี Silhouette Score	0.215
ครัวเรือนที่มีภาระหนี้สิน ซึ่งแบ่งกลุ่มด้วยวิธีดีปีสแกน โดยจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 1	0.800

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางจะพบว่าในกลุ่มครวเรือนที่มีภาระหนี้สิน วิธีดีปีสแกนให้ค่า Silhouette Coefficient ที่มากที่สุดคือ 0.800 จึงแสดงว่าในการแบ่งกลุ่มด้วยวิธีดีปีสแกนที่ eps มีค่าเท่ากับ 5.5, MinPts มีค่าเท่ากับ 19 แบ่งกลุ่มได้เหมาะสมที่สุด แต่เนื่องจากครวเรือนมีหนี้สินแบ่งกลุ่มด้วยวิธีดีปีสแกน ได้เพียง 1 กลุ่ม จึงสรุปได้ว่าครวเรือนมีหนี้สินไม่สามารถแบ่งกลุ่มด้วยวิธีดีปีสแกนได้ จึงเลือกการแบ่งกลุ่มด้วยวิธีเคมีน ที่ K=2 ที่มีค่า Silhouette Coefficient รองลงมา คือ 0.215

ตารางที่ 4.12 การเปรียบเทียบการแบ่งกลุ่มของครวเรือนที่ไม่มีภาระหนี้สิน

วิธีการแบ่งกลุ่ม	Silhouette Coefficient
ครวเรือนที่ไม่มีภาระหนี้สิน ซึ่งแบ่งกลุ่มด้วยวิธีเคมีน โดยหาจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2 ด้วยวิธี Silhouette Score	0.350
ครวเรือนที่ไม่มีภาระหนี้สิน ซึ่งแบ่งกลุ่มด้วยวิธีดีปีสแกน โดยจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2	0.763

จากตารางจะพบว่าในกลุ่มของครวเรือนที่ไม่มีภาระหนี้สิน วิธีดีปีสแกนให้ค่า Silhouette Coefficient ที่มากที่สุดคือ 0.763 จึงแสดงว่าในการแบ่งกลุ่มด้วยวิธีดีปีสแกนที่ eps มีค่าเท่ากับ 5.5, MinPts มีค่าเท่ากับ 19 แบ่งกลุ่มได้เหมาะสมที่สุด แต่เนื่องจากร้อยละของจำนวนตัวอย่างในกลุ่มที่ 1 มีค่าเท่ากับ 0.12 และในกลุ่มที่ 2 มีค่าเท่ากับ 99.88 ซึ่งมีค่าแตกต่างกันมากเกินไป จึงสรุปได้ว่าวิธีดีปีสแกน ไม่สามารถแบ่งกลุ่มครวเรือนที่ไม่มีหนี้สินได้อย่างเหมาะสม จึงเลือกการแบ่งกลุ่มด้วยวิธีเคมีน ที่ K=2 ที่มีค่า Silhouette Coefficient รองลงมา คือ 0.350

4.5 การทดสอบความแตกต่างของค่าเฉลี่ย

กลุ่มครวเรือนที่มีภาระหนี้สิน

จากข้อมูลกลุ่มครวเรือนที่มีภาระหนี้สิน ได้วิธีการแบ่งกลุ่มที่เหมาะสมที่สุดคือวิธีเคมีน โดยหาจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2 ด้วยวิธี Silhouette Score และทำการเปรียบเทียบค่าเฉลี่ยของตัวแปรในแต่ละกลุ่มได้ดังตาราง 4.13

ตารางที่ 4.13 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมีนคร้วเรือนที่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2

ชื่อตัวแปร	Cluster 1	Cluster 2	P-value
จำนวนสมาชิกของครัวเรือน	4.93	2.33	0.000*
อายุหัวหน้าครัวเรือน	56.17	53.68	0.000*
จำนวนผู้พิการในครัวเรือน	0.16	0.08	0.000*
จำนวนผู้ที่กำลังศึกษาในครัวเรือน	1.44	0.26	0.000*
จำนวนสวัสดิการด้านการรักษาพยาบาล	5.31	2.43	0.000*
จำนวนโครงการช่วยเหลือจากรัฐบาล	4.55	2.00	0.000*
จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล	0.50	0.41	0.000*
จำนวนผู้หารายได้ของครัวเรือน	2.51	1.65	0.000*
จำนวนของผู้สูงอายุและเด็กในครัวเรือน	1.74	0.36	0.000*
(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	137,969.62	119,484.45	0.012*
(หนี้ในระบบ)จำนวนหนี้สินของการศึกษา	7,565.12	6,634.98	0.369
(หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	195,163.94	135,843.66	0.000*
(หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	73,294.25	30,100.99	0.000*
(หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร	91,779.42	47,170.94	0.000*
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	3,794.54	2,490.97	0.557
(หนี้นอกระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	887.50	794.62	0.820
(หนี้นอกระบบ)จำนวนหนี้สินของการศึกษา	85.86	117.58	0.608
(หนี้นอกระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	3,334.52	2,409.13	0.080
(หนี้นอกระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	2,525.63	1,376.20	0.187
(หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร	1,186.20	467.88	0.133
(หนี้นอกระบบ)จำนวนหนี้สินอื่นๆ	111.72	353.58	0.043*
จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย	1,322.85	1,267.89	0.446
จำนวนเงินชำระหนี้การศึกษา	79.89	90.68	0.470
จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ	4,758.47	3,398.00	0.000*
จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร	1,117.11	518.25	0.000*

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการศึกษาเท่านั้น ไม่สามารถนำไปเผยแพร่โดยไม่ได้รับอนุญาต

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้คัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	Cluster 1	Cluster 2	P-value
จำนวนเงินชำระหนี้การทำการเกษตร	1,283.70	733.16	0.000*
จำนวนเงินชำระหนี้อื่นๆ	25.09	23.06	0.816
จำนวนเงินที่ซื้อที่อยู่อาศัย	67.76	33.51	0.426
จำนวนรายได้เฉลี่ยต่อเดือน	42,051.72	28,420.23	0.000*
จำนวนรายจ่ายเฉลี่ยต่อเดือน	24,308.40	16,679.11	0.000*

หมายเหตุ : ค่าสถิติทดสอบและค่าองศาอิสระแสดงในตารางที่ 1 ในภาคผนวก ค.

จากตารางที่ 4.13 ผลลัพธ์ที่ได้จากการทดสอบสมมติฐานของค่าเฉลี่ยประชากร 2 กลุ่ม ที่ระดับนัยสำคัญ 0.05 พบว่า ตัวแปรที่มีค่า P-value เท่ากับ 0.000 ซึ่งน้อยกว่าระดับนัยสำคัญ 0.05 ได้แก่ จำนวนสมาชิกของครัวเรือน อายุหัวหน้าครัวเรือน จำนวนผู้พิการในครัวเรือน จำนวนผู้ที่กำลังศึกษาในครัวเรือน จำนวนสวัสดิการด้านการรักษาพยาบาล จำนวนโครงการช่วยเหลือจากรัฐบาล จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล จำนวนผู้หารายได้ของครัวเรือน จำนวนของผู้สูงอายุและเด็กในครัวเรือน (หนี้ในระบบ) จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ (หนี้ในระบบ) จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร (หนี้ในระบบ) จำนวนหนี้สินของการทำการเกษตร จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตรจำนวน เงินชำระหนี้การทำการเกษตร จำนวนรายได้เฉลี่ยต่อเดือน และจำนวนรายจ่ายเฉลี่ยต่อเดือน แสดงว่า ตัวแปรที่ได้กล่าวไปข้างต้นทั้งหมดใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่มีภาระหนี้สินแตกต่างกันที่ระดับนัยสำคัญ 0.05

นอกจากนี้ตัวแปรที่มีค่า P-value น้อยกว่าระดับนัยสำคัญ 0.05 และมีค่า P-value มากกว่า 0.000 ได้แก่ (หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย และ(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ แสดงว่า ตัวแปร(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย และ(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ ใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่มีภาระหนี้สินแตกต่างกันที่ระดับนัยสำคัญ 0.05

ในขณะที่ตัวแปรที่มีค่า P-value มากกว่าระดับนัยสำคัญ 0.05 ได้แก่ (หนี้ในระบบ)จำนวนหนี้สินของการศึกษา (หนี้ในระบบ)จำนวนหนี้สินอื่นๆ (หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย (หนี้ในระบบ)จำนวนหนี้สินของการศึกษา (หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ (หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร (หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย จำนวนเงินชำระหนี้การศึกษา จำนวนเงินชำระหนี้อื่นๆ และจำนวนเงินที่ซื้อที่อยู่อาศัย แสดงว่า ตัวแปรที่ได้กล่าวไปข้างต้นทั้งหมดใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่มีภาระหนี้สินไม่แตกต่างกันที่ระดับนัยสำคัญ 0.05

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน

จากข้อมูลกลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน ได้ทำการแบ่งกลุ่มด้วยวิธีเคมีนโดยหาจำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2 ด้วยวิธี Silhouette Score ซึ่งสามารถแบ่งกลุ่มได้เหมาะสมที่สุด และได้ทำการเปรียบเทียบค่าเฉลี่ยของตัวแปรในแต่ละกลุ่มที่ได้ดังตาราง 4.14

ตารางที่ 4.14 ค่าเฉลี่ยตัวแปรจากการแบ่งกลุ่มวิธีเคมีนครัวเรือนที่ไม่มีภาระหนี้สินที่จำนวนกลุ่มที่เหมาะสม (K) เท่ากับ 2

ชื่อตัวแปร	Cluster 1	Cluster 2	P-value
จำนวนสมาชิกของครัวเรือน	1.70	4.19	0.000*
อายุหัวหน้าครัวเรือน	57.93	57.98	0.848
จำนวนผู้พิการในครัวเรือน	0.09	0.15	0.000*
จำนวนผู้ที่กำลังศึกษาในครัวเรือน	0.09	1.13	0.000*
จำนวนสวัสดิการด้านการรักษาพยาบาล	1.72	4.46	0.000*
จำนวนโครงการช่วยเหลือจากรัฐบาล	1.62	3.68	0.000*
จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล	0.00	0.01	0.000*
จำนวนผู้หารายได้ของครัวเรือน	0.94	1.95	0.000*
จำนวนของผู้สูงอายุและเด็กในครัวเรือน	0.44	1.49	0.000*
(หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	49.32	477.48	0.068
(หนี้ในระบบ)จำนวนหนี้สินของการศึกษา	26.39	291.78	0.006*
(หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	508.40	2,011.49	0.001*
(หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	81.87	183.68	0.305
(หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร	110.66	467.00	0.173
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	49.50	180.18	0.478
(หนี้นอกระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย	0.00	0.72	0.317
(หนี้นอกระบบ)จำนวนหนี้สินของการศึกษา	0.00	15.32	0.289
(หนี้นอกระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ	40.01	94.86	0.121
(หนี้นอกระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร	18.53	16.22	0.871
(หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร	31.26	13.51	0.423

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ของเจ้าของเนื้อหา หากมีข้อผิดพลาดประการใด ผู้เขียนขออภัยไว้ ณ ที่นี้ ไม่ขอรับผิดชอบใดๆ หากมีการนำเนื้อหาไปใช้
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชื่อตัวแปร	Cluster 1	Cluster 2	P-value
(หนี้ในระบบ)จำนวนหนี้สินอื่นๆ	15.19	0.00	0.301
จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัย	3.01	18.83	0.213
จำนวนเงินชำระหนี้การศึกษา	0.22	0.57	0.413
จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ	10.52	63.59	0.000*
จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร	1.31	1.28	0.979
จำนวนเงินชำระหนี้ทำการเกษตร	3.48	19.29	0.004*
จำนวนเงินชำระหนี้อื่นๆ	0.00	0.00	0.258
จำนวนเงินที่ซื้อที่อยู่อาศัย	2.80	0.00	0.000*
จำนวนรายได้เฉลี่ยต่อเดือน	18,330.69	36,348.17	0.000*
จำนวนรายจ่ายเฉลี่ยต่อเดือน	10,158.71	18,973.61	0.000*

หมายเหตุ : ค่าสถิติทดสอบและค่าองศาอิสระแสดงในตารางที่ 2 ในภาคผนวก ค.

จากตารางที่ 4.14 ผลลัพธ์ที่ได้จากการทดสอบสมมติฐานของค่าเฉลี่ยประชากร 2 กลุ่ม ที่ระดับนัยสำคัญ 0.05 พบว่า ตัวแปรที่มีค่า P-value เท่ากับ 0.000 ซึ่งน้อยกว่าระดับนัยสำคัญ 0.05 ได้แก่ จำนวนสมาชิกของครัวเรือน จำนวนผู้พิการในครัวเรือน จำนวนผู้ที่กำลังศึกษาในครัวเรือน จำนวนสวัสดิการด้านการรักษาพยาบาล จำนวนโครงการช่วยเหลือจากรัฐบาล จำนวนกองทุนกู้ยืมที่ได้จากรัฐบาล จำนวนผู้หารายได้ของครัวเรือน จำนวนของผู้สูงอายุและเด็กในครัวเรือน จำนวนเงินชำระหนี้การอุปโภคบริโภคอื่นๆ จำนวนเงินที่ซื้อที่อยู่อาศัย จำนวนรายได้เฉลี่ยต่อเดือน และจำนวนรายจ่ายเฉลี่ยต่อเดือน แสดงว่า ตัวแปรที่ได้กล่าวไปข้างต้นทั้งหมดใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินแตกต่างกันที่ระดับนัยสำคัญ 0.05

นอกจากนี้ตัวแปรที่มีค่า P-value น้อยกว่าระดับนัยสำคัญ 0.05 และมีค่า P-value มากกว่า 0.000 ได้แก่ (หนี้ในระบบ)จำนวนหนี้สินของการศึกษา (หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ และจำนวนเงินชำระหนี้ทำการเกษตร แสดงว่า ตัวแปรทั้ง 3 ตัวนี้ใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินแตกต่างกันที่ระดับนัยสำคัญ 0.05

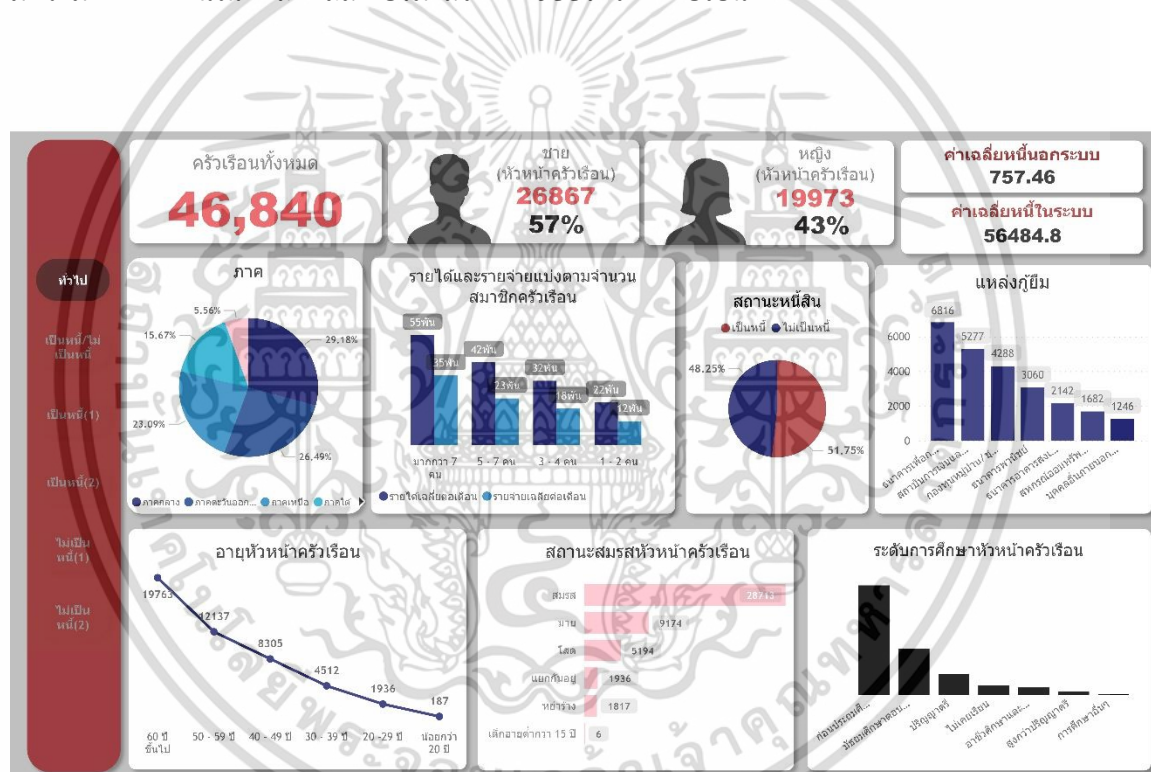
ในขณะที่ตัวแปรที่มีค่า P-value มากกว่าระดับนัยสำคัญ 0.05 ได้แก่ อายุหัวหน้าครัวเรือน (หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย (หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่ทางการเกษตร (หนี้ในระบบ)จำนวนหนี้สินของการทำการเกษตร (หนี้ในระบบ)จำนวนหนี้สินอื่นๆ (หนี้ในระบบ)จำนวนหนี้สินของการซื้อ/เช่าที่อยู่อาศัย (หนี้ในระบบ)จำนวนหนี้สินของการศึกษา (หนี้ในระบบ)จำนวนหนี้สินของการอุปโภคบริโภคอื่นๆ (หนี้ในระบบ)จำนวนหนี้สินของการทำธุรกิจที่ไม่ใช่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทางการเกษตร (หนี้นอกระบบ)จำนวนหนี้สินของการทำการเกษตร (หนี้นอกระบบ)จำนวนหนี้สินอื่นๆ จำนวนเงินชำระหนี้การซื้อ/เช่า ที่อยู่อาศัยจำนวนเงินชำระหนี้การศึกษา จำนวนเงินชำระหนี้การทำธุรกิจที่ไม่ใช่การเกษตร และจำนวนเงินชำระหนี้อื่นๆ แสดงว่า ตัวแปรที่ได้กล่าวไปข้างต้นทั้งหมดใน Cluster ที่ 1 และ Cluster ที่ 2 ของกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินไม่แตกต่างกันที่ระดับนัยสำคัญ 0.05

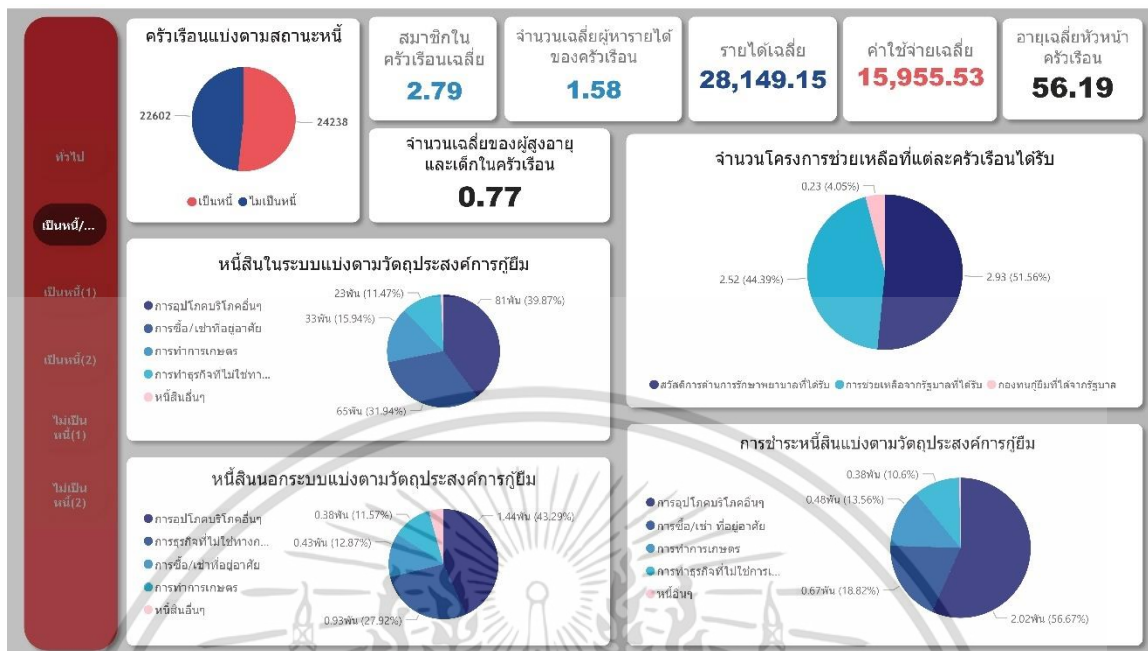
4.6 แดชบอร์ด

จากการทดสอบความแตกต่างของค่าเฉลี่ยผู้วิจัยนำตัวแปรจากการทดสอบที่ให้ผลการทดสอบว่ามีความแตกต่างกันในแต่ละคลัสเตอร์มาสร้างแดชบอร์ดได้ดังต่อไปนี้

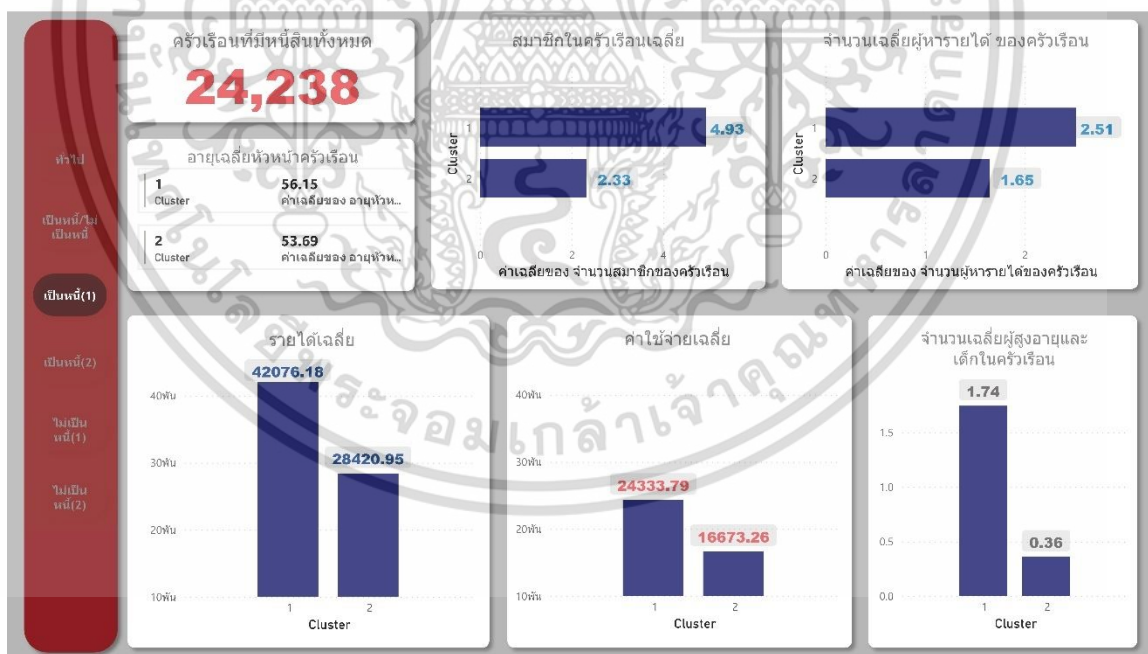


รูปที่ 4.5 แดชบอร์ดแสดงลักษณะทั่วไปของครัวเรือนไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

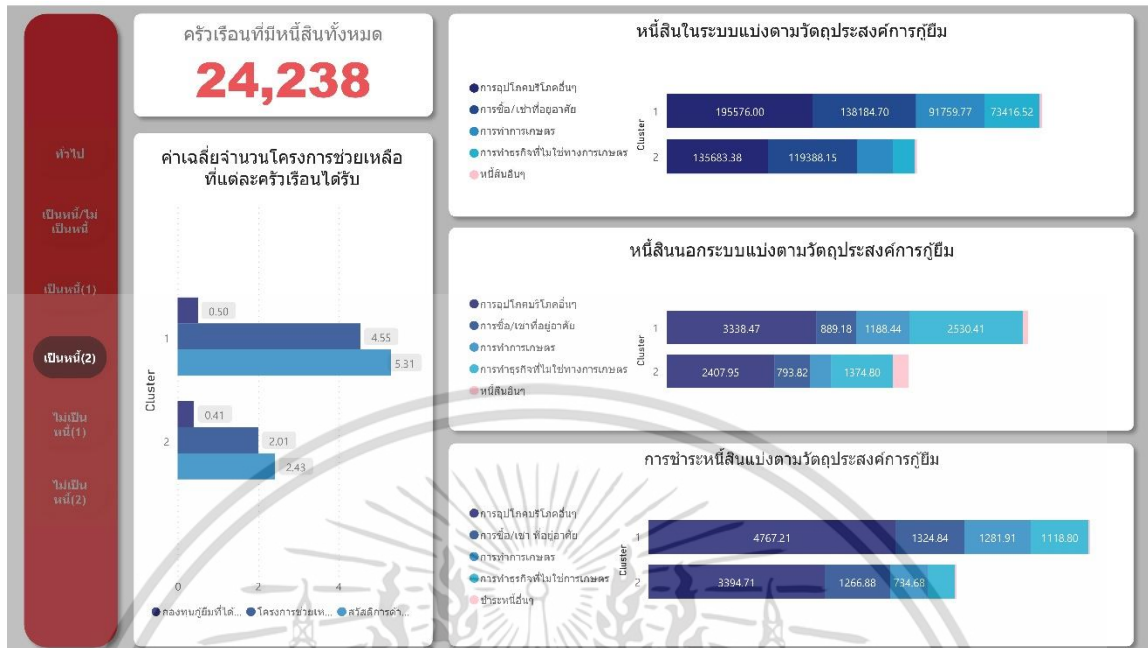


รูปที่ 4.6 แดชบอร์ดแสดงลักษณะโดยรวมของครัวเรือนไทยที่มีภาระหนี้สินและไม่มีภาระหนี้สิน

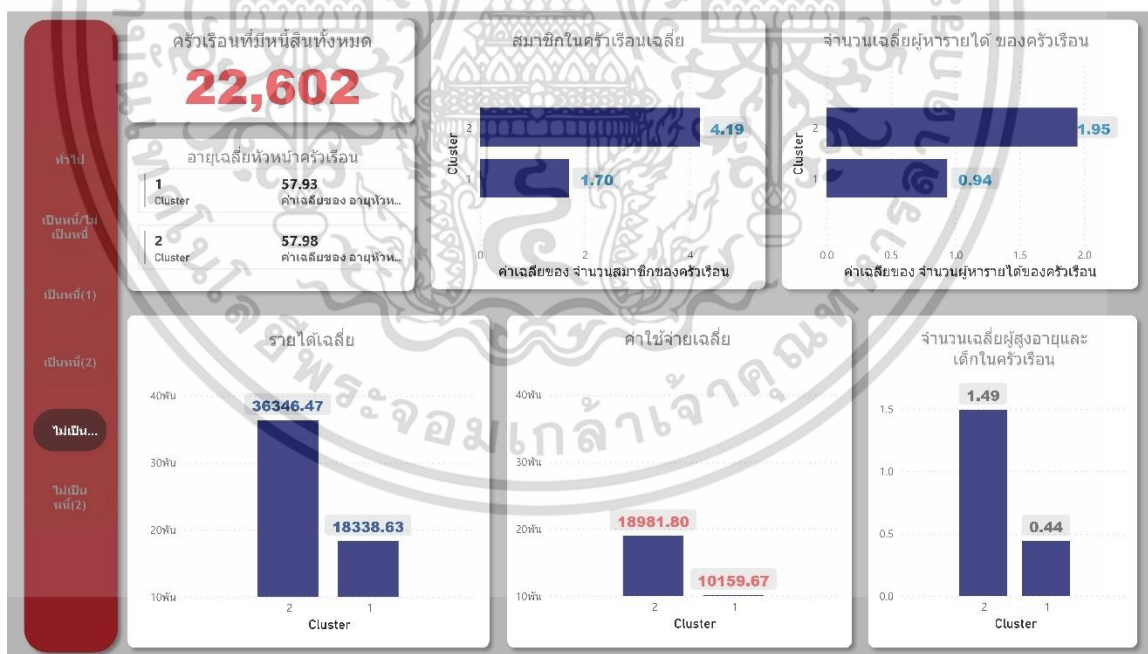


รูปที่ 4.7 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่มีภาระหนี้สินหน้า(1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีกร้นำไปใช้

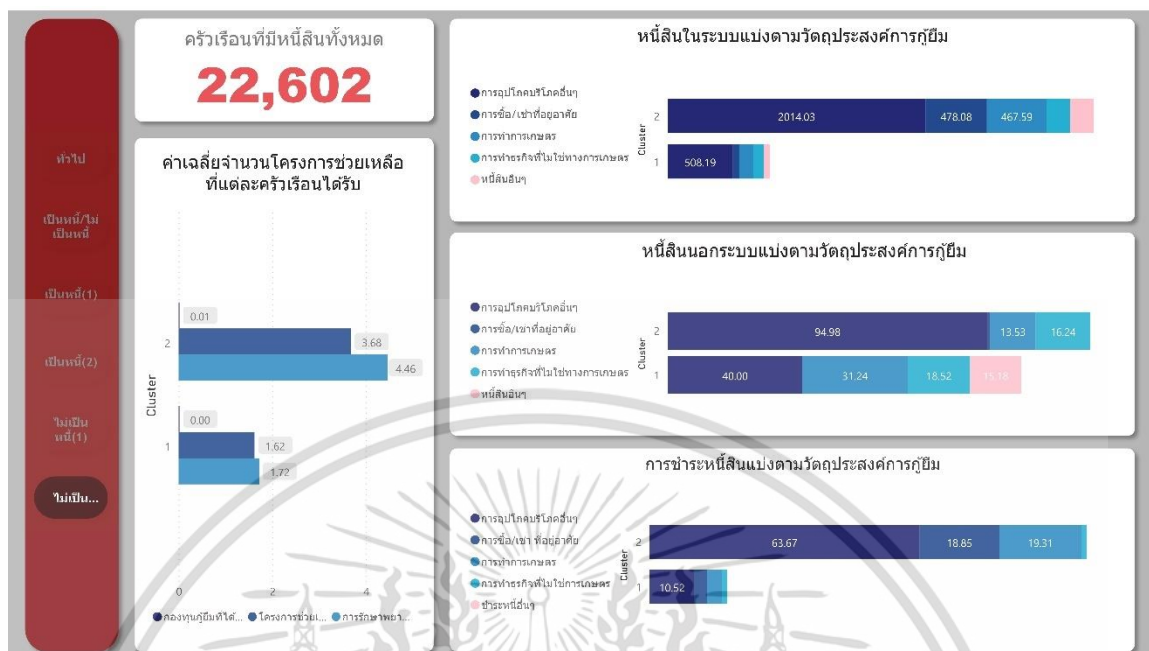


รูปที่ 4.8 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่ไม่มีภาระหนี้สินหน้า(2)



รูปที่ 4.9 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่ไม่มีภาระหนี้สินหน้า(1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 แดชบอร์ดแสดงลักษณะของครัวเรือนไทยที่ไม่มีภาระหนี้สินหน้า(2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การศึกษามีวัตถุประสงค์เพื่อเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของคนไทยโดยใช้วิธีเคมีนและดีปิสแกน เพื่อออกแบบแดชบอร์ดสำหรับใช้ในการนำเสนอข้อมูลภาระหนี้สินของคนไทย โดยนำข้อมูลทุติยภูมิจากสำนักงานสถิติแห่งชาติ เกี่ยวกับข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2564 จากแบบสำรวจครัวเรือนส่วนบุคคล ทั้งในเขตและนอกเขตเทศบาลทั่วประเทศ จำนวน 46,840 ครัวเรือน ในเดือนมกราคม - ธันวาคม 2564 และเมื่อพิจารณาตามลักษณะทั่วไปของประชากรไทยพบว่าครัวเรือนส่วนใหญ่กระจายอยู่ในภาคกลางบริเวณในเขตเทศบาล ส่วนใหญ่จะมีสมาชิกในครัวเรือน 1 ถึง 2 คน และมีจำนวนผู้ทำงานหารรายได้ 2 ถึง 3 คน มีสมาชิกอยู่ในวัยเด็กและวัยสูงอายุอย่างน้อย 1 คน และมีสมาชิกเป็นผู้พิการอย่างน้อย 1 คน ครัวเรือนมีรายได้ทั้งสิ้นเฉลี่ยเดือนละ 28,149 บาท ซึ่งส่วนใหญ่มีรายได้ทั้งรายได้ทั้งสิ้นอยู่ในช่วง 15,001 ถึง 30,000 บาทต่อเดือน ในส่วนของค่าใช้จ่ายทั้งสิ้นเฉลี่ยต่อเดือนของครัวเรือน พบว่าครัวเรือนมีค่าใช้จ่ายทั้งสิ้นเฉลี่ยต่อเดือนละ 15,956 บาท ซึ่งส่วนใหญ่มีค่าใช้จ่ายทั้งสิ้นอยู่ในช่วง 15,001 ถึง 30,000 บาทต่อเดือน โดยผู้วิจัยได้นำข้อมูลข้อมูลการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2564 มาต่อยอดโดยการทำเหมืองข้อมูล (Data Mining) ด้วยวิธีการแบ่งกลุ่มแบบเคมีนและดีปิสแกน โดยพิจารณาการเปรียบเทียบการแบ่งกลุ่มด้วย Silhouette coefficient

จากการศึกษาข้อมูลของกลุ่มครัวเรือนทั้งหมด 2 กลุ่ม ได้แก่

- กลุ่มครัวเรือนที่มีหนี้สิน
- กลุ่มครัวเรือนที่ไม่มีหนี้สิน

ในการแบ่งกลุ่มครัวเรือนที่มีภาระหนี้สินด้วยวิธีเคมีน ผลการวิจัย พบว่าสามารถแบ่งได้ 2 กลุ่ม โดยที่กลุ่มที่ 1 ประชากรมีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 5 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50 ถึง 59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 6 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 3 โครงการ และไม่มีการกู้ยืมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 3 คน มีผู้พิการและเด็กในครัวเรือน 2 คน และมีผู้ที่กำลังศึกษา 1 คน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 82,802 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 51,861 บาท ครัวเรือนมีวัตถุประสงค์ในการกู้ยืมที่มีจำนวนสูงสุด คือการกู้ยืมอื่นๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหาย โดยมีค่าเฉลี่ยอยู่ที่ 12,000,000 บาท และได้ทำการชำระหนี้ในส่วนของการเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กั๊ยมอื่น ๆ เช่น หนี้จากการค้าประกัน หนี้ค่าปรับ/จ่ายค่าเสียหายอยู่ที่ 33,600 บาทต่อเดือน และในกลุ่มที่ 2 ประชากรมีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 2 คน ค่าเฉลี่ยของอายุหัวหน้าครัวเรือนอยู่ในช่วง 50 ถึง 59 ปี ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 2 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 2 โครงการ และไม่มีการกั๊ยมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 2 คน ไม่มีผู้พิการ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 28,420 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 16,679 บาท ครัวเรือนมีวัตถุประสงค์ในการกั๊ยมที่มีจำนวนสูงสุด คือการกั๊ยมเพื่อใช้จ่ายอุปโภคบริโภคอื่น ๆ ในครัวเรือน โดยมีค่าเฉลี่ยอยู่ที่ 119,484 บาท และได้มีการชำระหนี้ในส่วนของการกั๊ยมเพื่อใช้จ่ายอุปโภคบริโภคอื่น ๆ ในครัวเรือนอยู่ที่ 3,398 บาทต่อเดือน และในการแบ่งกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินด้วยวิธีเคมีน ผลการวิจัย พบว่าสามารถแบ่งได้ 2 กลุ่ม โดยที่กลุ่มที่ 1 ประชากรมีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 2 คน ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 2 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 2 โครงการ และไม่มีการกั๊ยมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 1 คน ไม่มีผู้พิการ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 18,331 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 10,159 บาท และในกลุ่มที่ 2 ประชากรมีค่าเฉลี่ยของจำนวนสมาชิกในครัวเรือน 4 คน ค่าเฉลี่ยของจำนวนสวัสดิการด้านการรักษาพยาบาลที่ครัวเรือนได้รับอยู่ที่ 4 สวัสดิการ ได้รับโครงการช่วยเหลือของรัฐ 4 โครงการ และไม่มีการกั๊ยมเงินจากกองทุนต่างๆของรัฐ จำนวนเฉลี่ยของผู้หารายได้ในครัวเรือนอยู่ที่ 2 คน มีผู้พิการและเด็กในครัวเรือน 1 คน และมีผู้ที่กำลังศึกษา 1 คน มีรายได้ของครัวเรือนเฉลี่ยอยู่ที่ 36,348 บาท และมีค่าใช้จ่ายของครัวเรือนเฉลี่ย 18,974 บาท

5.2 อภิปรายผล

ปัญหาพิเศษเล่มนี้ได้ทำการศึกษาเกี่ยวกับการเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของประชากรไทยโดยวิธีเคมีนและดีปีสแกน โดยทำการแยกพิจารณาออกเป็น 2 กลุ่ม คือ กลุ่มครัวเรือนที่มีภาระหนี้สินและกลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน เพื่อศึกษาลักษณะทั่วไปและทำการแบ่งกลุ่มของครัวเรือนตามภาระหนี้สิน พบว่าในกลุ่มของครัวเรือนที่มีภาระหนี้สินและในกลุ่มของครัวเรือนที่ไม่มีภาระหนี้สิน มีลักษณะดังนี้

ประชากรของกลุ่มครัวเรือนที่มีภาระหนี้สินในกลุ่มที่ 1 เป็นครอบครัวขนาดใหญ่ มีจำนวนสมาชิกเฉลี่ยอยู่ที่ 5 คน โดยประชากรกลุ่มนี้มีผู้หารายได้ของครัวเรือนประมาณ 3 คน ทำให้มีรายได้ต่อครัวเรือนค่อนข้างสูง(42,051 บาท) และมีรายจ่ายสูงเช่นกัน(24,308 บาท) อาจเนื่องมาจากภายในครัวเรือนมีสมาชิกที่เป็นผู้สูงอายุและเด็กเฉลี่ยอยู่ที่ 2 คนต่อครัวเรือน และมีผู้ที่กำลังศึกษาในครัวเรือน 1 คน หนี้สิน

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของครัวเรือนกลุ่มนี้ส่วนใหญ่เป็นหนี้เนื่องจากการอุปโภคบริโภคอื่นๆ และประชากรกลุ่มนี้ได้รับสวัสดิการด้านการรักษาพยาบาลและโครงการช่วยเหลือจากรัฐบาลจำนวนค่อนข้างมาก

ประชากรของกลุ่มครัวเรือนที่มีภาระหนี้สินในกลุ่มที่ 2 เป็นครอบครัวที่มีขนาดเล็ก มีจำนวนสมาชิกประมาณ 2 คน โดยมีผู้หารายได้ของครัวเรือนเฉลี่ยอยู่ที่ 2 คน ประชากรกลุ่มนี้มีรายได้น้อยกว่ากลุ่มที่ 1 โดยมีรายได้ไม่ค่อนสูงมากนัก(28,420 บาท) และมีรายจ่ายที่ต่ำกว่ากลุ่มที่ 1 (16,679 บาท) อาจเนื่องมาจากส่วนใหญ่ไม่มีสมาชิกที่เป็นผู้สูงอายุ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน หนี้สินของครัวเรือนกลุ่มนี้ส่วนใหญ่เป็นหนี้เนื่องจากการอุปโภคบริโภคเช่นเดียวกันกับกลุ่มที่ 1 แต่ได้รับสวัสดิการด้านการรักษาพยาบาลและโครงการช่วยเหลือจากรัฐบาลน้อยกว่ากลุ่มที่ 1

ประชากรของกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินในกลุ่มที่ 1 เป็นครอบครัวขนาดเล็ก มีจำนวนสมาชิกประมาณ 2 คน ซึ่งประชากรกลุ่มนี้มีผู้หารายได้ของครัวเรือนเฉลี่ยอยู่ที่ 1 คน ไม่มีสมาชิกในครัวเรือนที่เป็นผู้สูงอายุ เด็ก และผู้ที่กำลังศึกษาในครัวเรือน โดยที่ส่วนใหญ่ครัวเรือนกลุ่มนี้จะมีรายได้ไม่สูงมากนัก(18,330 บาท) และมีรายจ่ายค่อนข้างน้อย(10,158 บาท) จึงทำให้มีเงินคงเหลือต่อเดือนไม่มากนัก(8,172 บาท) และประชากรกลุ่มนี้ได้รับสวัสดิการด้านการรักษาและโครงการช่วยเหลือจากภาครัฐจำนวนค่อนข้างน้อย

ประชากรของกลุ่มครัวเรือนที่ไม่มีภาระหนี้สินในกลุ่มที่ 2 เป็นครอบครัวขนาดกลาง มีจำนวนสมาชิกประมาณ 4 คน ซึ่งมีผู้หารายได้ของครัวเรือนเฉลี่ยอยู่ที่ 2 คน มีสมาชิกในครัวเรือนที่เป็นผู้สูงอายุ เด็ก หรือผู้ที่กำลังศึกษาในครัวเรือน โดยส่วนใหญ่ครัวเรือนกลุ่มนี้จะมีรายได้สูง(36,348 บาท) มีรายจ่ายที่สูงกว่ากลุ่มที่ 1(18,973 บาท) แต่มีเงินคงเหลือมากกว่ากลุ่มที่ 1(17,375 บาท) และประชากรกลุ่มนี้ได้รับจำนวนสวัสดิการด้านการรักษาและโครงการช่วยเหลือจากรัฐบาลค่อนข้างมากกว่ากลุ่มที่ 1

5.3 ข้อเสนอแนะ

เนื่องจากการศึกษานี้เป็นการวิเคราะห์ลักษณะทั่วไปของข้อมูลที่ได้ทำการแบ่งกลุ่ม หากสามารถต่อยอดด้วยการใช้เทคนิค Logistic Regression, Random Forest, Neural Network โดยให้การมีภาระหนี้สินเป็นตัวแปรตาม จะสามารถสร้างแบบจำลองเพื่อจำแนกครัวเรือนตามภาระหนี้สินได้

บรรณานุกรม

- กัลยา วานิชย์บัญชา. (2552). การวิเคราะห์ข้อมูลหลายตัวแปร. กรุงเทพมหานคร : ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย. พิมพ์ครั้งที่ 4.
- ชฎารัตน์ พิพัฒน์นันท์ และ สิริธร เจริญรัตน์. (2559). การวิเคราะห์ข้อมูลทางเศรษฐกิจและสังคมของครัวเรือนด้วยวิธีการแบ่งกลุ่มและหาความสัมพันธ์ สำหรับการทำให้เมืองข้อมูล. วารสารพัฒนาบริหารศาสตร์. 56(3), 138 – 161.
- มาริสสา สกุลวัฒนา. (2564). การศึกษาปัจจัยที่มีผลต่อการก่อหนี้กรณีศึกษาของกลุ่มพนักงานวัยเริ่มทำงานของบริษัทเอกชน. [ออนไลน์]. เข้าถึงได้จาก : <http://irithesis.swu.ac.th/dspace/bitstream/123456789/1721/1/gs621130245.pdf>
- รชต ตั้งนาราชชกิจ และ พิรญาณ์ รณภาพ. (2562). “หนี้ครัวเรือน : ปัญหาที่ทุกคนต้องร่วมด้วยช่วยกันแก้” ธนาคารแห่งประเทศไทย. [ออนไลน์]. เข้าถึงได้จาก : <https://www.bot.or.th/Thai/BOTMagazine/Pages/25650157TheknowledgeHouseholdDebt.aspx>
- รัตนารณณ์ เขาวลิตตระกูล. (2551). เอกสารประกอบการสอน เรื่อง ความรู้ทั่วไปเกี่ยวกับหนี้สิน. [ออนไลน์]. เข้าถึงได้จาก : http://www.pbntc.moe.go.th/inventor_teacher/datas/A%201.doc
- ราชบัณฑิตยสถาน. (2525). พจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2525. กรุงเทพฯ: อักษรเจริญทัศน์
- รุจิรา จุลภักดิ์, จีรภา เพชรพัฒนานันท์ และ ศิริเรือง พัฒน์ช่วย. (2564). การวิเคราะห์พยากรณ์หนี้สินจากบัญชีครัวเรือนเกษตรกรด้วยเทคนิคโครงข่ายประสาทเทียม. วารสารวิชาการการประยุกต์ใช้เทคโนโลยีสารสนเทศ, 7(1), 30 – 31.
- รศ.ดร.ชัชวาลย์ ศิลปกิจ. (2564). สถิติเบื้องต้น. [ออนไลน์]. เข้าถึงได้จาก : <http://med.mahidol.ac.th/psych/sites/default/files/public/pdf/Conference/Researchconference/57/Basic%20Statistics.pdf>
- วนิษา แผลงรักษา และ นิเวศ จิระวิจิตรชัย (2565). “การแบ่งกลุ่มลูกค้าโดยใช้เทคนิคการทำคลัสเตอร์แบบเคมีน”. วารสารวิชาการชายันท์เทศ, 3(2). 1-10.
- สรสา ชื่นโชคสันต์ และคณะ. (2562). “8 ข้อเท็จจริง ปัญหาการเงินของครัวเรือนไทย”. ธนาคารแห่งประเทศไทย. [ออนไลน์]. เข้าถึงได้จาก : https://www.bot.or.th/Thai/ResearchAndPublications/articles/Pages/Article_30Oct2019.aspx

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- สุกฤตา สงวนพันธ์. (2558). **Bankers’ Talk Household debt** (ฝ่ายสายนโยบายสถาบันการเงิน). *ธนาคารแห่งประเทศไทย, 3(1), 1-2.*
- โสมรัตน์ จันทรัตน์ และคณะ. (2562). "เข้าใจพลวัตหนี้ครัวเรือนไทยผ่าน Big Data ของเครดิตบูโร". aBRIDGEd.
- สำนักงานสถิติแห่งชาติ. (2565). หนี้สินของครัวเรือน. [ออนไลน์]. เข้าถึงได้จาก : https://catalog.nso.go.th/dataset/os_08_00011
- สำนักงานสถิติ. (2562). หนี้สินของครัวเรือนเกษตร ปี พ.ศ. 2562. [ออนไลน์]. เข้าถึงได้จาก : http://service.nso.go.th/nso/nsopublish/pubs/e-book/Agri_Household_Deb62/index.html
- หนึ่งหทัย ชัยอากร. (2559). การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้า ไม่นิ่ง.[ออนไลน์]. เข้าถึงได้จาก : <https://erp.mju.ac.th/acticleDetail.aspx?qid=551#:~:text=Data%20Mining%20เป็นเทคนิคใน.Data%20Mining%20แล้วจะได้>
- อาเดรียโน โด. อมาเรลา, วิลาวรรณย์ พึ่งตัว และ เต็ง จิงหยู. (2563). The Impact of the COVID-19 Pandemic on Household Debt in Thailand ผลกระทบจากการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) ที่มีต่อหนี้ครัวเรือนไทย. วารสารวิชาการมหาวิทยาลัยอีสเทิร์นเอเชีย. *12(1), 70-81.*
- อารีรัตน์ ลุนลลาด และ อมรรรัตน์ อภินันท์มกุล. (2564). ภาวะหนี้สินของครัวเรือนเกษตรทำนาเปรียบเทียบกับหนี้สินของครัวเรือนที่ประกอบอาชีพอื่น. พัฒนาการเศรษฐกิจปริทรรศน์. *16(1), 104-129.*
- Ajitesh Kumar. (2021). KMeans Silhouette Score Python Example. [online]. from : <https://vitalflux.com/elbow-method-silhouette-score-which-better/>
- BENMOUIZA Khalil and CHEKNANE Ali (2016). “Density-based spatial clustering of application with noise algorithm for the classification of solar radiation time series”. Proceeding of 2016 8th International Conference on Modelling, Identification and Control: ICMIC-2016. Algiers : Algeria. *8th, 279 - 283.*
- Eugenia Anello. (2022). “Top 9 Power BI Dashboard Examples”. [online]. from: <https://www.datacamp.com/blog/9-power-bi-dashboard-examples>
- Lalita Lowphansiriku. (2018). Clustering — DBSCAN คืออะไร. [online]. from : <https://medium.com/@lalita/clustering-dbscan-%E0%B8%84z%E0%B8%B7%E0%B8%AD%E0%B8%AD%E0%B8%B0%E0%B9%84%E0%B8%A3-116b5d5c9873>
- M. Q. Duong, L. H. Lam, Bui Thi Minh Tu, Giap Quang Huy, N. H. Hieu. (2019).

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

"A Combination of K-Mean Clustering and Elbow Technique in Mitigating Losses of Distribution Network". *GMSARN International Journal*. 13, 153 - 185.

Shu-Zhi Liu , Rashmi Sharan Sinha และ Seung-Hoon Hwang. (2021). "Clustering-Based Noise Elimination Scheme for Data Pre-Processing for Deep Learning Classifier in Fingerprint Indoor Positioning System". *Sensors* 2021. 21(13). from : <https://doi.org/10.3390/s21134349>



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก. คำสั่งที่ใช้ในการวิเคราะห์การแบ่งกลุ่มภาระหนี้สินของครัวเรือนประชากรไทย

ตารางที่ 1 คำสั่งเตรียมข้อมูลเพื่อวิเคราะห์การแบ่งกลุ่มภาระหนี้สินของครัวเรือนประชากรไทย

```
#นำข้อมูลเข้าไลบรารีที่จำเป็นในการเตรียมข้อมูล
import pandas as pd
import numpy as np

#นำเข้าข้อมูลที่เตรียมไว้สำหรับการวิเคราะห์
Household_df = pd.read_csv('/Users/HP/Desktop/NoDebt(Newdata).csv')

#การทำข้อมูลสูญหาย
Microdata_df.isnull().sum()

#การแปลงข้อมูลให้อยู่ในมาตรฐานเดียวกันด้วย StandardScaler
from sklearn.preprocessing import StandardScaler
Household_scaler = StandardScaler()

Household_sc = Household_scaler.fit_transform(Household_df)
#Convert to table format – StandardScaler
Cols = Household_df.columns
Household_sc = pd.DataFrame(Household_sc, columns=[Cols])
```

ตารางที่ 2 คำสั่งที่ใช้ในการวิเคราะห์หาค่า K ที่เหมาะสมที่สุด

```
#การหาค่า K ที่เหมาะสมที่สุดด้วยวิธี Silhouetted Score
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

def sil_score(Household_sc, from_k=2, to_k=7):
    sils=[]
    for k in range(from_k, to_k+1):
```

เอกสารนี้เป็นเอกสารสงวนลิขสิทธิ์สำหรับใช้ในการเรียนการสอนเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ทางการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

model = KMeans(n_clusters=k,init='kmeans++',
               random_state=42).fit(Household_sc)
silhouette_avg = silhouette_score(Household_sc, model.labels_).round(4)
sils.append([silhouette_avg,k])
return sils

ss = sil_score(Household_sc, 2, 7)
print(f'scores ={ss}')
print(f'optimal number of clusters ={max(ss)[1]}')

#คำสั่งแสดงผลกราฟ Silhouetted
from yellowbrick.cluster import SilhouetteVisualizer
K = (2,3,4,5,6,7)
for k in K:
    model = KMeans(n_clusters = k,init='kmeans++',random_state = 42)
    Visualizer = SilhouetteVisualizer(model)
    Visualizer.fit(Household_sc)
    Visualizer.poof()

```

ตารางที่ 3 คำสั่งที่ใช้ในการวิเคราะห์หาค่า eps และ Mintpts ที่เหมาะสมที่สุด

```

from sklearn.neighbors import NearestNeighbors
neigh = NearestNeighbors(n_neighbors=20)
nbrs = neigh.fit(Household_sc)
distances, indices = nbrs.kneighbors(Household_sc)

#คำสั่งแสดงผลกราฟ k-distance
from matplotlib import pyplot as plt
distances = np.sort(distances, axis=0)
distances = distances[:,1]

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

plt.figure(figsize=(10,8))
plt.title('NearestNeighbors')
plt.plot(distances)

#ประมาณค่าจากจุดหักคอกจากกราฟและกำหนดค่าเพื่อหาค่า eps และ MinPts
from sklearn.cluster import DBSCAN
min_samples = range(10,20)
eps = np.arange(2.5,6,1.5)

output = []
for ms in min_samples:
    for ep in eps:
        labels = DBSCAN(min_samples = ms, eps = ep).fit(Household_sc).labels_
        output.append((ms, ep))
min_samples, eps = sorted(output, key=lambda x:x[-1])[-1]
print(f"min_samples: {min_samples}")
print(f"eps: {eps}")

```

ตารางที่ 4 คำสั่งที่ใช้ในการแบ่งกลุ่ม

```

#คำสั่งการแบ่งกลุ่มวิธีเคมีน ใส่น cluster จากค่า K ที่ได้จาก Silhouetted
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters = .., init='kmeans++', random_state=42).fit(Household_sc)
identified_cluster1 = kmeans.predict(Household_sc)

#คำสั่งการแบ่งกลุ่มวิธีเคมีน ใส่นค่า min_samples และ eps ที่ได้จาก Nearest Neighbors
labels = DBSCAN(min_samples=.., eps = 5.5).fit(Household_sc).labels_
clusters = len(Counter(labels))
print(f"Number of clusters: {clusters}")
print(f"Number of outliers: {Counter(labels)[-1]}")

```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4 คำสั่งที่ใช้ในการเปรียบเทียบ

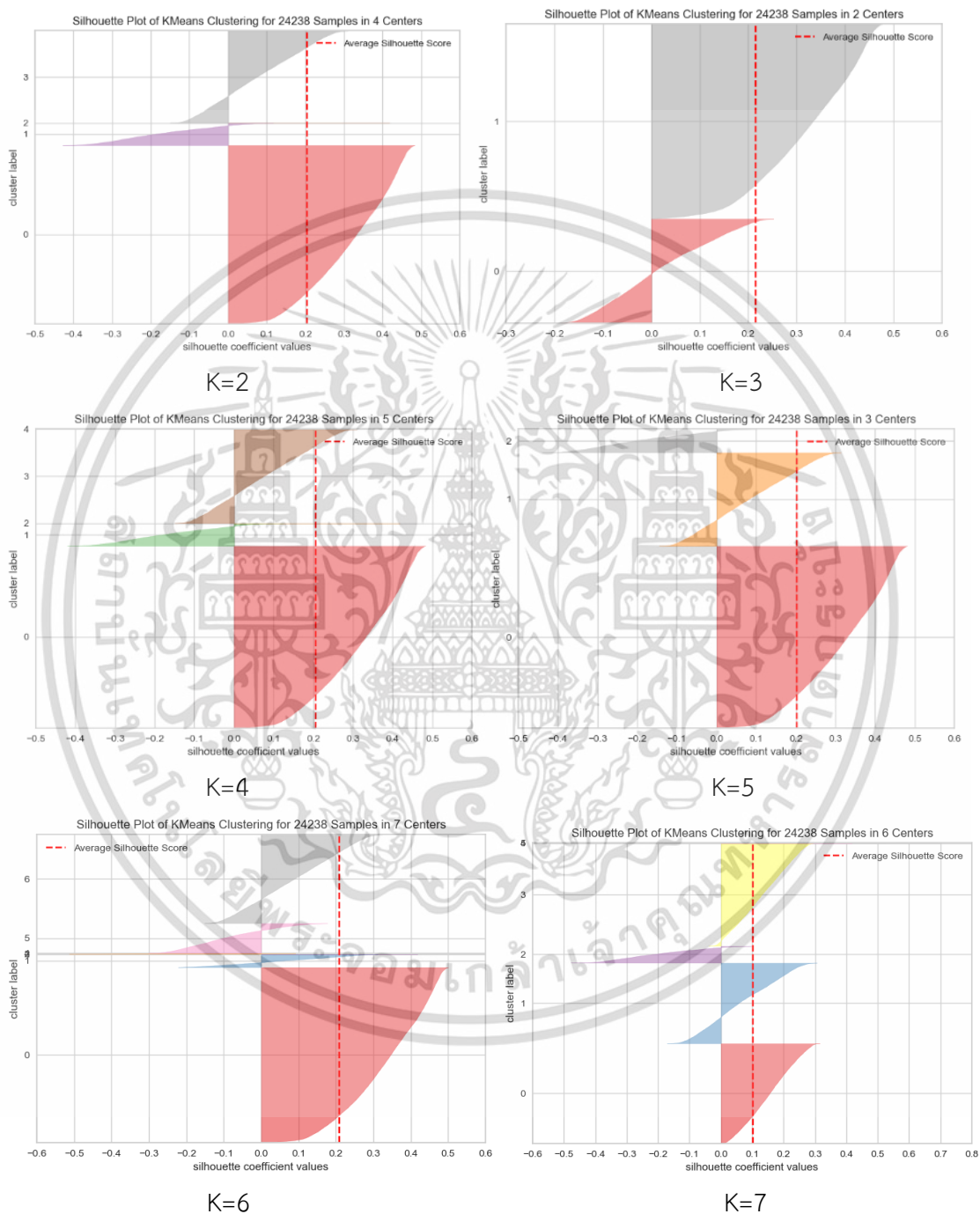
```
#คำสั่งเปรียบเทียบโดยใช้ Silhouette Coefficient
from sklearn import metrics
print(f"Silhouette Coefficient of Kmeans (K=..): {metrics.silhouette_score(Household_sc,
identified_cluster):.3f}")
print(f"Silhouette Coefficient DBSCAN: {metrics.silhouette_score(Household_sc,
labels):.3f}")
```



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

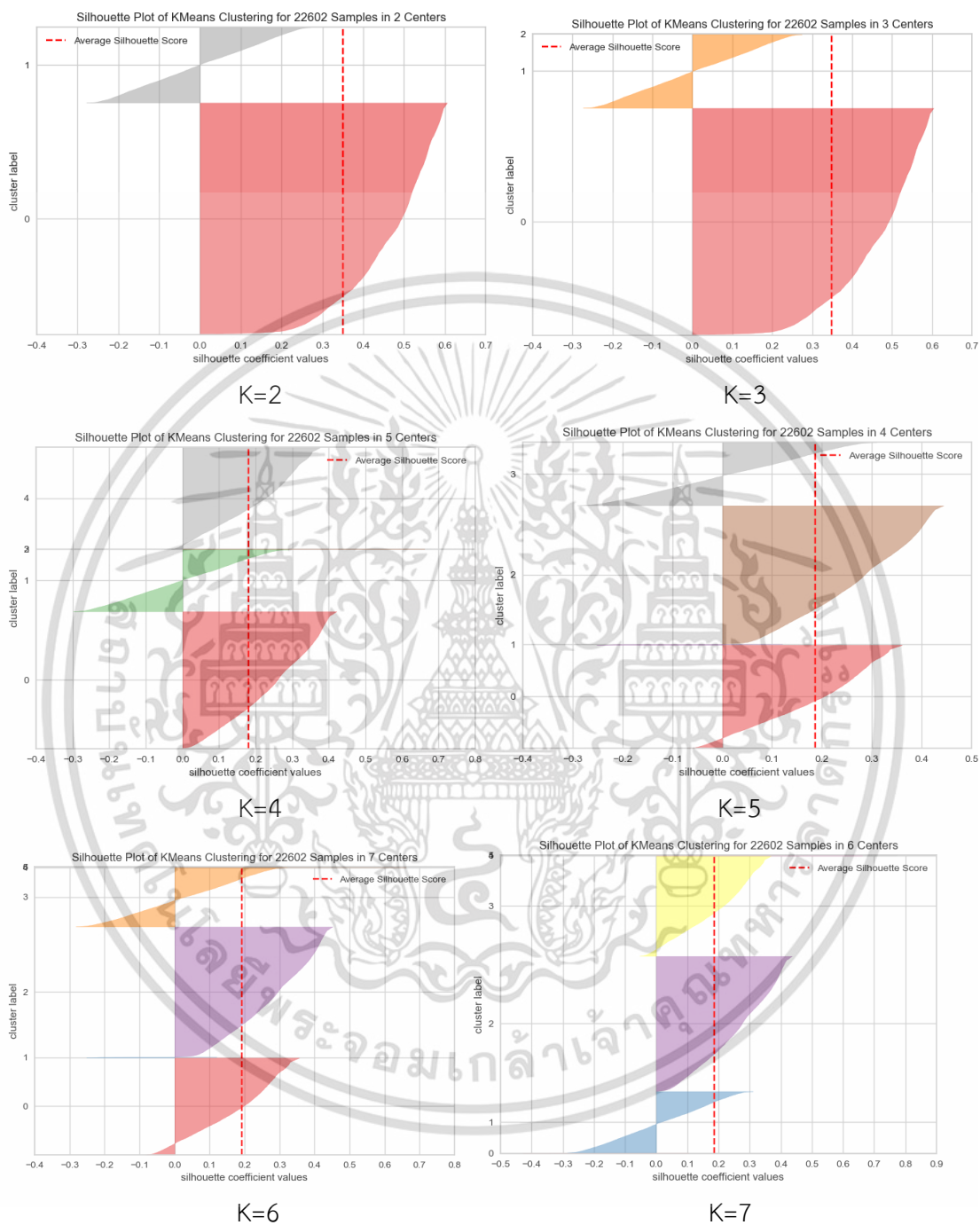
ภาคผนวก ข.กราฟการหาค่า K ที่เหมาะสมที่สุดด้วยวิธี Silhouette Score

กลุ่มครัวเรือนที่มีภาระหนี้สิน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กลุ่มครัวเรือนที่ไม่มีภาระหนี้สิน



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค. การทดสอบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม

ตารางที่ 1 การทดสอบสมมติฐานของค่าเฉลี่ยกลุ่มক্রว่เรือนที่มีภาวะหนี้สิน 2 กลุ่ม ที่ระดับนัยสำคัญ 0.05

		Independent Samples Test						
		HM01_Members		HM04_Age		HM07_PWD		
		Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		518.032		49.444		955.812	
	Sig.		0.000		0.000		0.000	
t-test for Equality of Means	t		189.794	169.495	13.934	13.795	16.204	14.001
	df		24236	12684.469	24236	16768.285	24236	11707.219
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000
	Mean Difference		2.598	2.598	2.460	2.460	0.087	0.087
	Std. Error Difference		0.014	0.015	0.177	0.178	0.005	0.006
	95% Confidence Interval of the Difference	Lower		2.572	2.568	2.114	2.110	0.076
Upper			2.625	2.628	2.806	2.809	0.097	0.099

Independent Samples Test									
			HM14_Education		HM22_Welfare		HM29_AidGov		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		5586.825		1412.317		1488.320		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		131.235	109.838	175.949	152.297	97.713	87.139	
	df		24236	10891.445	24236	11756.668	24236	12639.797	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	
	Mean Difference		1.184	1.184	2.883	2.883	2.542	2.542	
	Std. Error Difference		0.009	0.011	0.016	0.019	0.026	0.029	
	95% Confidence Interval of the Difference	Lower		1.166	1.162	2.851	2.846	2.491	2.484
		Upper		1.201	1.205	2.915	2.920	2.592	2.599

Independent Samples Test									
			HM36_LoansGov		HM41_Earner		HM42_Child_Elder		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		286.218		946.034		3088.029		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		12.052	11.541	72.124	66.099	134.227	113.566	
	df		24236	15269.718	24236	13555.014	24236	11156.230	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	
	Mean Difference		0.095	0.095	0.867	0.867	1.384	1.384	
	Std. Error Difference		0.008	0.008	0.012	0.013	0.010	0.012	
	95% Confidence Interval of the Difference	Lower		0.079	0.079	0.844	0.842	1.364	1.360
		Upper		0.110	0.111	0.891	0.893	1.404	1.408

Independent Samples Test									
			AD12_Land_FM		AD13_Education_FM		AD14_Consumer_FM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		27.910		3.878		181.186		
	Sig.		0.000		0.049		0.000		
t-test for Equality of Means	t		2.663	2.510	1.040	0.899	11.305	10.183	
	df		24236	14615.368	24236	11706.133	24236	12963.892	
	Sig. (2-tailed)		0.008	0.012	0.298	0.369	0.000	0.000	
	Mean Difference		18772.625498	18772.625498	949.870	949.870	59897.211550	59897.211550	
	Std. Error Difference		7049.291994	7478.737460	913.093	1056.830	5298.470223	5882.039492	
	95% Confidence Interval of the Difference	Lower		4955.577039	4113.355428	-839.850	-1121.693	49511.882089	48367.549532
		Upper		32589.673957	33431.895568	2739.589	3021.432	70282.541011	71426.873568

Independent Samples Test									
			AD15_Business_FM		AD16_Farm_FM		AD17_Other_FM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		75.536		244.517		1.732		
	Sig.		0.000		0.000		0.188		
t-test for Equality of Means	t		4.704	3.576	10.477	8.113	0.660	0.588	
	df		24236	9012.367	24236	9316.833	24236	12583.863	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.509	0.557	
	Mean Difference		43326.571646	43326.571646	44519.274318	44519.274318	1312.672076	1312.672076	
	Std. Error Difference		9210.344125	12115.699262	4249.217214	5487.139813	1988.494302	2233.740718	
	95% Confidence Interval of the Difference	Lower		25273.727303	19577.047881	36190.545673	33763.280582	-2584.899788	-3065.800421
		Upper		61379.415989	67076.095411	52848.002964	55275.268055	5210.243939	5691.144572

Independent Samples Test									
			AD18_Land_NFM		AD19_Education_NFM		AD20_Consumer_NFM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		0.207		0.807		15.114		
	Sig.		0.649		0.369		0.000		
t-test for Equality of Means	t		0.228	0.227	-0.449	-0.512	1.998	1.750	
	df		24236	16995.893	24236	23617.233	24236	12086.737	
	Sig. (2-tailed)		0.819	0.820	0.653	0.608	0.046	0.080	
	Mean Difference		95.203	95.203	-31.449	-31.449	929.973	929.973	
	Std. Error Difference		416.670	418.909	70.011	61.369	465.438	531.510	
	95% Confidence Interval of the Difference	Lower		-721.496	-725.903	-168.674	-151.737	17.686	-111.871
		Upper		911.902	916.309	105.776	88.838	1842.261	1971.818

Independent Samples Test									
			AD21_Business_NFM		AD22_Farm_NFM		AD23_Other_NFM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		9.706		14.966		9.841		
	Sig.		0.002		0.000		0.002		
t-test for Equality of Means	t		1.571	1.319	1.949	1.501	-1.569	-2.023	
	df		24236	10981.018	24236	9226.963	24236	20838.707	
	Sig. (2-tailed)		0.116	0.187	0.051	0.133	0.117	0.043	
	Mean Difference		1155.226	1155.226	720.863	720.863	-241.318	-241.318	
	Std. Error Difference		735.577	875.606	369.879	480.181	153.774	119.311	
	95% Confidence Interval of the Difference	Lower		-286.550	-561.119	-4.122	-220.397	-542.725	-475.177
		Upper		2597.002	2871.571	1445.848	1662.124	60.089	-7.459

Independent Samples Test									
			AE02_PD_Land		AE03_PD_Education		AE04_PD_Consumer		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		2.738		2.273		146.141		
	Sig.		0.098		0.132		0.000		
t-test for Equality of Means	t		0.768	0.762	-0.764	-0.723	11.286	9.379	
	df		24236	16879.133	24236	14812.206	24236	10722.328	
	Sig. (2-tailed)		0.443	0.446	0.445	0.470	0.000	0.000	
	Mean Difference		57.717	57.717	-10.563	-10.563	1372.996	1372.996	
	Std. Error Difference		75.191	75.779	13.832	14.604	121.650	146.398	
	95% Confidence Interval of the Difference	Lower		-89.663	-90.817	-37.675	-39.188	1134.555	1086.029
		Upper		205.097	206.251	16.550	18.062	1611.438	1659.963

Independent Samples Test									
			AE05_PD_Business		AE06_PD_Farm		AE07_PD_Other		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		85.988		220.664		0.266		
	Sig.		0.000		0.000		0.606		
t-test for Equality of Means	t		5.061	3.870	9.698	7.784	0.256	0.233	
	df		24236	9105.064	24236	9976.857	24236	13299.725	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.798	0.816	
	Mean Difference		600.681	600.681	547.040	547.040	2.097	2.097	
	Std. Error Difference		118.693	155.233	56.408	70.282	8.179	8.990	
	95% Confidence Interval of the Difference	Lower		368.036	296.389	436.477	409.273	-13.934	-15.524
		Upper		833.327	904.973	657.603	684.807	18.128	19.718

Independent Samples Test									
			AE09_Amount_BuyHouse		Total_Income		Total_Cost		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		3.131		431.676		511.276		
	Sig.		0.077		0.000		0.000		
t-test for Equality of Means	t		0.886	0.797	29.846	24.765	34.525	29.952	
	df		24236	12922.057	24236	10688.918	24236	11819.712	
	Sig. (2-tailed)		0.376	0.426	0.000	0.000	0.000	0.000	
	Mean Difference		34.409	34.409	13659.2269505083	13659.2269505083	7666.6055	7666.6055	
	Std. Error Difference		38.856	43.190	457.6600977948	551.5624911608	222.0600	255.9651	
	95% Confidence Interval of the Difference	Lower		-41.750	-50.250	12762.1848427293	12578.0619066128	7231.3541	7164.8717
		Upper		110.569	119.069	14556.2690582872	14740.3919944038	8101.8569	8168.3393

ตารางที่ 2 การทดสอบสมมติฐานของค่าเฉลี่ยกลุ่มক্রวเรือนที่ไม่มีภาวะหนี้สิน 2 กลุ่ม ที่ระดับนัยสำคัญ 0.05

Independent Samples Test									
		HM01_Members		HM04_Age		HM07_PWD			
		Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed		
Levene's Test for Equality of Variances	F		1364.039		123.807		363.776		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		-191.504	-149.797	-0.179	-0.192	-9.730	-8.089	
	df		22600	6876.669	22600	10687.009	22600	7317.493	
	Sig. (2-tailed)		0.000	0.000	0.858	0.848	0.000	0.000	
	Mean Difference		-2.491	-2.491	-0.045	-0.045	-0.054	-0.054	
	Std. Error Difference		0.013	0.017	0.254	0.236	0.006	0.007	
	95% Confidence Interval of the Difference	Lower		-2.517	-2.524	-0.543	-0.509	-0.065	-0.068
		Upper		-2.466	-2.458	0.452	0.418	-0.044	-0.041

Independent Samples Test									
		HM14_Education		HM22_Welfare		HM29_AidGov			
		Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed		
Levene's Test for Equality of Variances	F		9896.790		2404.862		1573.136		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		-131.183	-84.276	-181.044	-134.513	-79.292	-64.083	
	df		22600	5911.184	22600	6566.332	22600	7102.079	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	
	Mean Difference		-1.044	-1.044	-2.745	-2.745	-2.059	-2.059	
	Std. Error Difference		0.008	0.012	0.015	0.020	0.026	0.032	
	95% Confidence Interval of the Difference	Lower		-1.059	-1.068	-2.774	-2.785	-2.109	-2.122
		Upper		-1.028	-1.019	-2.715	-2.705	-2.008	-1.996

Independent Samples Test									
			HM36_LoansGov		HM41_Earner		HM42_Child_Elder		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		185.672		297.643		2619.914		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		-6.809	-4.501	-75.172	-64.948	-92.889	-72.895	
	df		22600	6018.531	22600	7646.496	22600	6898.020	
	Sig. (2-tailed)		0.000	0.000	0.000	0.000	0.000	0.000	
	Mean Difference		-0.006	-0.006	-1.010	-1.010	-1.053	-1.053	
	Std. Error Difference		0.001	0.001	0.013	0.016	0.011	0.014	
	95% Confidence Interval of the Difference	Lower		-0.008	-0.009	-1.037	-1.041	-1.075	-1.081
		Upper		-0.005	-0.004	-0.984	-0.980	-1.030	-1.024

Independent Samples Test									
			AD12_Land_FM		AD13_Education_FM		AD14_Consumer_FM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		36.612		66.493		79.246		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		-3.025	-1.823	-4.079	-2.769	-4.495	-3.463	
	df		22600	5700.035	22600	6127.941	22600	6779.763	
	Sig. (2-tailed)		0.002	0.068	0.000	0.006	0.000	0.001	
	Mean Difference		-428.781	-428.781	-265.773	-265.773	-1505.844	-1505.844	
	Std. Error Difference		141.757	235.147	65.162	95.976	335.029	434.863	
	95% Confidence Interval of the Difference	Lower		-706.634	-889.759	-393.496	-453.920	-2162.523	-2358.312
		Upper		-150.928	32.197	-138.050	-77.626	-849.165	-653.376

Independent Samples Test									
			AD15_Business_FM		AD16_Farm_FM		AD17_Other_FM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		6.182		19.858		4.569		
	Sig.		0.013		0.000		0.033		
t-test for Equality of Means	t		-1.245	-1.025	-2.235	-1.364	-1.069	-0.710	
	df		22600	7236.862	22600	5736.576	22600	6039.965	
	Sig. (2-tailed)		0.213	0.305	0.025	0.173	0.285	0.478	
	Mean Difference		-102.074	-102.074	-356.975	-356.975	-130.932	-130.932	
	Std. Error Difference		81.958	99.602	159.690	261.780	122.521	184.360	
	95% Confidence Interval of the Difference	Lower		-262.718	-297.322	-669.980	-870.162	-371.081	-492.343
		Upper		58.570	93.174	-43.971	156.211	109.217	230.479

Independent Samples Test									
			AD18_Land_NFM		AD19_Education_NFM		AD20_Consumer_NFM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		12.316		13.846		13.365		
	Sig.		0.000		0.000		0.000		
t-test for Equality of Means	t		-1.754	-1.000	-1.860	-1.060	-1.834	-1.549	
	df		22600	5542.000	22600	5542.000	22600	7448.394	
	Sig. (2-tailed)		0.079	0.317	0.063	0.289	0.067	0.121	
	Mean Difference		-0.722	-0.722	-15.335	-15.335	-54.988	-54.988	
	Std. Error Difference		0.411	0.722	8.243	14.461	29.988	35.501	
	95% Confidence Interval of the Difference	Lower		-1.528	-2.136	-31.491	-43.683	-113.767	-124.581
		Upper		0.085	0.693	0.821	13.014	3.791	14.605

Independent Samples Test									
			AD21_Business_NFM		AD22_Farm_NFM		AD23_Other_NFM		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		0.071		1.008		1.393		
	Sig.		0.790		0.315		0.238		
t-test for Equality of Means	t		0.133	0.162	0.502	0.800	0.590	1.035	
	df		22600	14246.685	22600	22288.513	22600	17058.000	
	Sig. (2-tailed)		0.894	0.871	0.616	0.423	0.555	0.301	
	Mean Difference		2.287	2.287	17.714	17.714	15.183	15.183	
	Std. Error Difference		17.218	14.078	35.287	22.130	25.726	14.664	
	95% Confidence Interval of the Difference	Lower		-31.460	-25.307	-51.452	-25.662	-35.243	-13.561
		Upper		36.035	29.882	86.879	61.090	65.608	43.927

Independent Samples Test									
			AE02_PD_Land		AE03_PD_Education		AE04_PD_Consumer		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		14.035		4.611		199.731		
	Sig.		0.000		0.032		0.000		
t-test for Equality of Means	t		-1.874	-1.244	-1.074	-0.819	-7.108	-4.457	
	df		22600	6035.126	22600	6716.341	22600	5826.361	
	Sig. (2-tailed)		0.061	0.213	0.283	0.413	0.000	0.000	
	Mean Difference		-15.845	-15.845	-0.349	-0.349	-53.155	-53.155	
	Std. Error Difference		8.454	12.736	0.325	0.427	7.478	11.925	
	95% Confidence Interval of the Difference	Lower		-32.415	-40.812	-0.987	-1.185	-67.812	-76.533
		Upper		0.725	9.122	0.288	0.487	-38.498	-29.777

Independent Samples Test									
			AE05_PD_Business		AE06_PD_Farm		AE09_Amount_BuyHouse		
			Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		0.001		72.877		1.667		
	Sig.		0.970		0.000		0.197		
t-test for Equality of Means	t		-0.018	0.027	-4.273	-2.846	0.645	1.132	
	df		22600	21650.921	22600	6048.865	22600	17058.000	
	Sig. (2-tailed)		0.985	0.979	0.000	0.004	0.519	0.258	
	Mean Difference		0.026	0.026	-15.830	-15.830	2.794	2.794	
	Std. Error Difference		1.449	0.987	3.705	5.562	4.329	2.468	
	95% Confidence Interval of the Difference	Lower		-2.813	-1.908	-23.092	-26.734	-5.691	-2.043
		Upper		2.866	1.961	-8.569	-4.927	11.280	7.631

Independent Samples Test						
		Total_Income			Total_Cost	
		Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	
Levene's Test for Equality of Variances	F		658.936		1210.301	
	Sig.		0.000		0.000	
t-test for Equality of Means	t		-32.757	-20.600	-51.211	-37.710
	df		22600	5836.211	22600	6517.232
	Sig. (2-tailed)		0.000	0.000	0.000	0.000
	Mean Difference		-18007.8327969462	-18007.8327969462	-8822.1358	-8822.1358
	Std. Error Difference		549.7326737326	874.1507534882	172.2690	233.9463
	95% Confidence Interval of the Difference	Lower		-19085.3467458285	-19721.4921833401	-9159.7949
Upper			-16930.3188480639	-16294.1734105523	-8484.4767	-8363.5243



งานทะเบียนคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

คำรับรองเล่มโครงการพิเศษ/ปัญหาพิเศษ/สหกิจศึกษา

วันที่ 8 เดือน มิถุนายน พ.ศ 2566

ข้าพเจ้า นางสาวไพรวาด แลซอ รหัสนักศึกษา 62050807
นางสาววิไลรัตน์ สายสะอาด รหัสนักศึกษา 62050834
นางสาวสิริทิพย์ ทองโต รหัสนักศึกษา 62050843

นักศึกษาหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชา สถิติประยุกต์ ภาควิชา สถิติ

ขอรับรองว่าโครงการพิเศษ/ปัญหาพิเศษ/สหกิจศึกษา เรื่อง

ชื่อภาษาไทย การเปรียบเทียบการแบ่งกลุ่มครัวเรือนตามภาระหนี้สินของประชากรไทยโดยวิธีเคมีนและดีปัสแกน

ชื่อภาษาอังกฤษ COMPARISON OF HOUSEHOLD SEGMENTATION BY DEBT OF THE THAI POPULATION
BY K-MEANS AND DBSCAN METHODS

ปีการศึกษา 2565

เป็นผลงานวิจัยที่มีได้คัดลอกหรือละเมิดลิขสิทธิ์ของผู้อื่นและได้ผ่านการตรวจสอบความซ้ำซ้อนเรียบร้อยแล้ว และได้
แนบเอกสารการตรวจสอบการลอกเลียนงานวรรณกรรมที่ตรวจสอบจากเล่มโครงการพิเศษ/ปัญหาพิเศษ/สหกิจศึกษา
ฉบับสมบูรณ์แล้ว

โปรแกรมอักขราวิสุทธิ์ 4.91 %

ลงชื่อ.....ไพรวาด.....

(ไพรวาด แลซอ)

นักศึกษา

ลงชื่อ.....วิไลรัตน์.....

(วิไลรัตน์ สายสะอาด)

นักศึกษา

ลงชื่อ.....สิริทิพย์.....

(สิริทิพย์ ทองโต)

นักศึกษา

ข้าพเจ้า ผศ.ดร.ยูวดี กล่อมวิเศษ อาจารย์ที่ปรึกษาปัญหาพิเศษ ได้ตรวจสอบปัญหาพิเศษของนักศึกษาข้างต้นแล้ว

ขอรับรองว่าเป็นผลงานวิจัยของนักศึกษาจริงและมีเนื้อหาสมบูรณ์ จึงลงชื่อไว้เป็นหลักฐาน

ลงชื่อ.....ยูวดี.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
(ผศ.ดร.ยูวดี กล่อมวิเศษ)

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อาจารย์ที่ปรึกษา